OPRE REPORT #2014-88

OUTCOMES OF PROJECT LAUNCH: Cross-Site Evaluation Findings, Volume II







OUTCOMES OF PROJECT LAUNCH:

CROSS-SITE EVALUATION FINDINGS, VOLUME II

OPRE Report # 2014-88

December 2014

Barbara Goodson, Todd Grindal, Catherine Darrow, Margaret Gwaltney, Deborah Klein Walker, Brandy Wyant, Cristofer Price, Abt Associates Inc.

Submitted to:
Laura Hoard, Project Officer
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Contract No. HHSP233200800591G

Project Director: Margaret Gwaltney
Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02138

The Cross-Site Evaluation of Project LAUNCH was funded by the Substance Abuse and Mental Health Services Administration, U.S. Department of Health and Human Services. This report and other reports sponsored by the Office of Planning, Research and Evaluation through an Interagency Agreement with the Substance Abuse and Mental Health Services Administration are available at http://www.acf.hhs.gov/programs/opre

OVERVIEW

Project LAUNCH (Linking Actions for Unmet Needs in Children's Health) is a federal grant program administered by the Substance Abuse and Mental Health Services Administration (SAMHSA) to promote the social-emotional, cognitive, physical and behavioral health of children from birth to eight years of age (SAMHSA, 2008). Grantees are funded to pursue goals of improving early childhood systems and services in pilot communities selected because of a high need for services for families and children with significant risk factors, insufficient services, and significant health and economic disparities.[1]

The findings of the Cross-Site Evaluation (CSE) of Project LAUNCH are presented in two volumes: Volume I presents results of the process evaluation and Volume II, the results of the outcome evaluation. The CSE covers findings through 2013 for the first three cohorts of LAUNCH grantees: 6 grantees in Cohort 1 (funded in 2008), 12 grantees in Cohort 2 (funded in 2009), and 6 grantees in Cohort 3 (funded in 2010).

The outcome evaluation uses meta-analysis to summarize effects on providers, parents and children, and local and state systems, from studies conducted and reported by grantee local evaluators. Across the local evaluations, outcome data were reported for a subset of the total services and systems change activities being implemented by the grantees. The CSE used the Rating System for Strong and Emerging Evidence Designs (R-SEED)® framework to rate the strength of the evidence from the local evaluations. Key findings include:

- The overall average effect of Project LAUNCH on providers was small-to-moderate (ES = 0.38). When effects are adjusted for strength of evidence, the average decreases (ES = 0.32) because of the preponderance of effects that received the lowest evidence ratings.
- The overall average effect on parent outcomes was moderate (ES = 0.48). When effects are adjusted for strength of evidence, the overall average is lower (ES = 0.44).
- The average effect on children was small (ES = 0.28). When effects are adjusted for strength of evidence, the overall average does not change.
- The average effects on local systems change was small (ES = 0.32), and the average is lower when the effects are adjusted for strength of evidence (ES = 0.27). The average effect on state systems change was moderate (ES = 0.50). This average also is smaller when effects are weighted by strength of evidence (ES = 0.41).

The outcome data for Project LAUNCH suggest that the providers, parents, and children who are part of the LAUNCH initiative are on positive trajectories, even though the fact that the local evaluation studies do not include comparison groups means that LAUNCH cannot claim to be the causal agent of this growth. Future evaluations should test whether growth in program participants

[1] In the pilot communities, LAUNCH grantees worked to increase access to screening, assessment, and referral to appropriate services for young children and their families; increase integration of mental and behavioral health in primary care and early childhood education settings; expand use of use of culturally-relevant, evidence-based prevention and wellness promotion practices; increase workforce knowledge of children's social and emotional development and preparation to deliver high quality care.

is greater than growth in similar families not in LAUNCH. The examination of potential predictors of variation in effects focused on two implementation indices—one related to the depth of the integration of behavioral health in individual LAUNCH programs, and the second related to the breadth of initiatives to improve the child service system at the local and state levels. These analyses generated findings that suggested that these indices were related to the size of the observed effects, which suggests that future analyses of outcomes could provide guidance about the most effective strategies for promoting LAUNCH goals.

EXECUTIVE SUMMARY

Background of Project LAUNCH

Project LAUNCH (Linking Actions for Unmet Needs in Children's Health) is a federal grant program administered by the Substance Abuse and Mental Health Services Administration (SAMHSA). The goal of the program is to promote the physical, social, emotional, cognitive, and behavioral health of young children from birth to eight years of age (SAMHSA, 2008). State/tribal-level grantees that receive funding from SAMHSA subsequently provide funding to a pilot community to pursue goals of improving early childhood systems and early childhood services. The goals of Project LAUNCH are to:

- Increase the integration of mental and behavioral health into early childhood services, including home visiting, family strengthening and parent education, early childhood education and care, and primary care;
- Expand the use of culturally-relevant, evidence-based prevention and wellness promotion practices (EBPs) for integration of behavioral health and implementation of mental health consultation in home visiting, family strengthening and parent education, early childhood education, and primary care;
- Increase access to screening, assessment, and referral to appropriate services for young children and their families in a range of child-serving settings; and
- Improve coordination and collaboration across local, state, tribal and federal agencies serving young children and their families.

This national program has three guiding principles. First, the program assumes a *holistic perspective* to health that encompasses the physical, social, emotional, cognitive, and behavioral health of all children from birth to age eight. Second, the program employs an *ecological framework*, giving attention to all settings that influence children's health and wellbeing: the family, home environment, school, pediatric care settings, neighborhood, and community. Third, the program takes a *public health approach*. It focuses on improving all systems that serve young children and their families and incorporates prevention and health promotion activities that encourage intervening early before problems occur.

As of September 2014, SAMHSA has funded five cohorts of grantees. This outcome report includes the first three cohorts, funded in 2008 (6 grantees), 2009 (12 grantees), and 2010 (6 grantees).

Cross-Site Evaluation Approach: Outcomes of Project LAUNCH

The Cross-Site Evaluation of the outcomes of Project LAUNCH is a meta-analysis of the outcomes of individual LAUNCH-supported programs across the 24 LAUNCH grantees in Cohorts 1-3. The Cross-Site Evaluation was guided by the following questions:

- How has health and well-being improved for young children (birth – 8 years) participating in LAUNCH-supported services?

- How has the local child services system changed during the Project LAUNCH grant implementation?
- How has the state child services system changed during the Project LAUNCH grant implementation?
- How have knowledge and practices changed for providers of LAUNCH-supported services?
- How have the health and well-being improved for parents of children participating in LAUNCH-supported services?

In addition to these key research questions, the analyses also explored a small set of hypotheses about implementation factors that might explain variation in the outcomes:

- Within strategies, are there implementation factors that are associated with variation in outcomes for providers? For parents or children?
- At the grantee-level, are there implementation factors that are associated with variation in the average effects of different LAUNCH projects for providers? For parents or children?

This outcomes report is the second volume of a two-volume report on the Cross-Site Evaluation findings. It includes the first three cohorts, funded in 2008 (6 grantees), 2009 (12 grantees), and 2010 (6 grantees) and presents findings for grantees when they were at different stages of implementation: grantees in Cohort 1 were completing their last grant year; Cohort 2 grantees were ending their fourth year; and grantees in Cohort 3 were completing their third year.

Volume I of this report presents findings about the implementation of Project LAUNCH and how child and family services in the community have been enhanced as a result of Project LAUNCH. Data reported by grantees in a Web-based data portal (Figure E-1), obtained from grantees' annual program reports, and collected by the Cross-Site Evaluation during key informant interviews are summarized in that report and provide a context for understanding the outcome data for each of the Project LAUNCH strategies. The report details how the LAUNCH initiative has introduced new services and enhanced existing services, with a focus on integration of behavioral health into the child and family services system. The emphasis of Project LAUNCH on behavioral health has enabled grant recipients not only to fill service gaps and enhance existing services with practices that attend to the social-emotional health of young children, but has also led to new efforts that develop the infrastructure within state and local governments and support evidence-based service delivery that meets the comprehensive needs of at-risk children and their families.

The data on outcomes that were used in the Cross-Site Evaluation meta-analysis come from studies conducted by the local LAUNCH evaluators and reported in annual end-of-year evaluation reports from grantees. The local evaluation studies measured outcomes for providers, parents, and/or children as well as local and state systems changes. The number of annual evaluation reports available varies by grantee cohort, with the most findings available from Cohort 1 grantees and the least from Cohort 3 grantees: Each Cohort 1 grantee had annual reports from five years of implementation, Cohort 2 grantees had reports from four years of implementation, and Cohort 3 grantees had reports on three years of implementation. The Cross-Site Evaluation also uses results

from a set of separately-funded "Special Studies" conducted by local evaluators, which compared community-wide child outcomes in the LAUNCH community to outcomes in one or more comparison communities. Six grantees from Cohorts 1 and 2 were awarded additional evaluation funds for these Special Studies, based on a competitive process. These grantees together implemented ten different studies looking at community-wide outcomes.

Figure E-1. Data Sources for Implementation Context

Cohort (Year initially funded)	Implementation Year 1	Implementation Year 2	Implementation Year 3	Implementation Year 4	Implementation Year 5
Cohort 1 (2008)	Web portal: 9 reporting periods (Fall 2009, Spring 2010, Fall 2010, Spring 2011, Fall 2011, Spring 2012, Fall 2012, Spring 2013, Fall 2013)				
Cohort 2 (2009)	Web portal: 7 reporting periods (Fall 2010, Spring 2011, Fall 2011, Spring 2012, Fall 2012, Spring 2013, Fall 2013)				
Cohort 3 (2010)	Web portal: 5 reporting periods (Fall 2011, Spring 2012, Fall 2012, Spring 2013, Fall 2013)				

Summary of Findings from the Cross-Site Evaluation of Project LAUNCH

For providers, parents and children, the overall average effects of LAUNCH are in a positive direction. In general, across the LAUNCH prevention and promotion strategies, the overall effects for child outcomes are smaller than the effects for parent, provider outcomes, and systems outcomes. The size of the effects for the different LAUNCH service strategies are not unlike the size of the effects reported in previous meta-analyses of similar services. The evidence for outcomes came predominantly from one-group designs and, as a result, the outcomes cannot be assumed to have been caused by Project LAUNCH. To summarize the findings:

- Across all program strategies and all strengths of evidence, the overall average effect on providers is small-to-moderate (ES = 0.38). The average effect ranges from a high of 0.55 for the effects assessed as providing Weak evidence to a low of 0.15 for the small number of effects rated as providing Strong evidence (Figure 2). This trend is consistent with other meta-analyses that report larger average effects for weaker studies.

  When the overall average effects are adjusted for strength of evidence, the average is decreased because of the large number of effects at the lowest evidence ratings that are "down-weighting" in the weighting system adopted for the meta-analysis. When the effects are adjusted for strength of evidence, the overall average effect of LAUNCH is smaller (ES = 0.32 versus 0.38 unweighted). The same pattern occurs for the average effects on providers, parents and children. The overall size of the average effects suggests that if the meta-analysis only included studies with more rigorous designs, there would likely still be an effect but it would be smaller.

- Across all program strategies, the overall average effect on parent outcomes was small-to-moderate (ES=0.48) when the effects were not weighted by strength of evidence. When the overall average is adjusted for strength of evidence, it is lower (ES = 0.44)
- Across all program strategies, the average effect on children was small (ES = 0.28). When the overall average is adjusted for strength of evidence, it is virtually identical (ES = 0.29).
- The average effects on local systems change was small (ES = 0.32). When the effects were adjusted for their strength of evidence, the average drops (ES = 0.27). The average effect on state systems change was moderate in size (ES = 0.50). This average also was smaller when effects were weighted by strength of evidence (ES = 0.41). All of the evidence on systems change was rated as providing Weak evidence or evidence that is Limited with Reservations. The size of the effects on systems change is very likely related to the fact that the studies contributing evidence were of low rigor and therefore more likely to have bigger effect sizes.
- On average, the majority of providers in each of the program strategies reported that their knowledge of children's socio-emotional development, of appropriate service options for children with behavioral concerns, and their use of mental health consultation had increased "some" or "substantially" since the time when LAUNCH became involved with the program. On average, parents were positive about the amount of help they received from the LAUNCH programs. Regardless of the type of program, a high rate of parents participating in the LAUNCH-supported home visiting programs felt that the program was helpful for their family, for their parenting skills, and for their child's development. A number of parent and provider outcomes were generated from the SAMHSA Provider Survey and the SAMHSA Parent Survey. Both are retrospective pre-post measures that produce self-reported scores on change as a result of LAUNCH involvement with their service, not actual change scores. These outcomes were not included in the meta-analysis. The fact that surveys produce change scores, not separate baseline or posttest scores, means that they do not meet R-SEED®[2] standards for strength of evidence, and we were not able to calculate the standardized effect sizes required for the outcome analyses.

Two measures of implementation were developed and used in relational analyses of the LAUNCH effects, to explore whether variation on effects might be related to differences in how grantees implemented their programs and services. One measure of implementation described, for each LAUNCH-supported program, the depth of the LAUNCH-supported activities to integrate child mental health into services. (Each program was given a point on each of the 12 possible integration activities.) When analyses correlated this implementation factor to overall effects, findings suggested that variation in program effects may, in fact, be related to variation in the kinds of supports that LAUNCH grantees have implemented and are implementing to bring mental and behavioral health into the child and family services systems.

A second implementation variable is defined as the number of categories of systems change in which grantees undertook initiatives in the LAUNCH community (local level) and at the state level. This variable is considered to measure the *breadth* of systems change activities—i.e.,

[2] Rating System for Strong and Emerging Evidence Designs (R-SEED)®.

implementation of systems change activities across different categories (or domains). It was hypothesized that greater breadth of systems change activities to improve state, tribal, and local child service systems will predict stronger outcomes for providers, parents and children. The analyses partially confirm this hypothesis: The variable created to measure breadth of systems change activities was positively related to the overall effects on parents and overall effects on the extent of collaboration reported at the state level (although not to effects on providers or children).

Additional analyses examined whether the average effect sizes varied by the grantee cohort. Where there were significant relationships, the average effects for Cohort 1 were higher, on average, than outcomes for the other two cohorts. Reasons why Cohort 1 programs would have better outcomes than were found in Cohorts 2 and 3 include the fact that the Cohort 1 programs have been in place longer and have had more time for program enhancements and/or programs to mature and become more effective and for parents to have received longer doses of a program's services. If these are reasons for the better performance of Cohort 1 programs, then over time, the effects in the later cohorts should "catch up" when their programs have been in place for more years.

There were a few positive relationships that suggested that the implementation of the LAUNCH model was related to the size of the observed effects. However, the few findings, laid against the total number of relationships tested, and laid against the concerns with the quality of the evidence, means that the observed results have to be seen as suggestive and potentially worth testing in the future as opposed to confirming true relationships.

Summary of Design and Methods for the Cross-Site Evaluation of the Outcomes of Project LAUNCH

For each annual evaluation report from a grantee, outcome data on LAUNCH-supported programs or systems change were identified and systematically coded for analysis. The coding included information on the LAUNCH-supported program on which the outcomes were measured, the characteristics of the outcome measures, the design of the outcome study, the analysis model, and the findings. Using the raw outcome data, standard statistical software[3] was used to transform each effect or contrast into a standardized effect size (Hedges' *g*). The resulting effect size represents the size of a treatment-comparison difference relative to the standard deviation of the outcome measure. The standardized effect sizes were weighted by the size of the samples that generated the effects.[4] The assumption is that the effects that were measured on larger samples were more accurate representations of the true effects. The result is that the effects measured with larger samples have more weight in the analysis.

During the coding of the outcome data, if a study design carried out as part of a local evaluation did not generate a contrast, these outcomes were not included. For example, in some evaluations, outcomes were reported only at the end of a program, with no baseline or pre-test data on the

[3] The software program used for the LAUNCH analysis to calculate standardized effect sizes was Comprehensive Meta-Analysis Version 2®.

[4] The weighted effect size was calculated by multiplying the effect size by the inverse of the effect size standard error squared.

same respondents or data on an external comparison group. These outcomes were not included in the meta-analysis, since there was no basis for calculating a difference or an effect.

Each contrast, in addition to being transformed into a standardized effect size, was assigned a rating that communicates the level of confidence with which a difference (e.g., treatment-comparison or pre-post) can be attributed to the LAUNCH intervention being tested, i.e., that the difference represents a valid measure of a LAUNCH program effect. This rating is labeled as the "strength of evidence" of the contrast, or the internal validity of the estimate of the effect. The most rigorous designs, such as randomized control trials, provide the strongest evidence of a treatment effect, because the randomization ensures that the only statistical difference between the two groups being compared is the treatment. Contrasts from studies using other designs, such as quasi-experimental designs or pre-post designs, receive lower strength of evidence ratings because of potential competing hypotheses that could explain the difference, such as time or development. Knowing the strength of evidence for contrasts in a meta-analysis such as this one for Project LAUNCH allows us to describe the extent to which the result of the meta-analysis can be seen as causal evidence of the effects of LAUNCH.

For the Cross-Site Evaluation, the system used to assign a rating of the strength of the evidence of Project LAUNCH effects was the Rating System for Strong and Emerging Evidence Designs (R-SEED)® (Goodson, Price, Darrow, et al., in development). R-SEED was selected[5] because it encompasses the full range of designs, from the most rigorous (randomized control trials) to designs that do not involve external comparison groups (e.g., pre-post studies). Once the R-SEED ratings were generated, the meta-analysis used (a) information on the LAUNCH program that was coded by the reviewers, (b) the standardized effect sizes calculated from the findings coded by the reviewers, and (c) the R-SEED rating generated by the computer program. The meta-analyses used only the effects that met R-SEED evidence standards, although some of the analyses focused on the effects with the higher R-SEED ratings.

The average effect sizes for these analyses were calculated using two different methods. The first method calculated average effect sizes using all of the relevant contrasts weighted by sample size, regardless of their R-SEED rating for strength of evidence. We also conducted analyses that took into account the strength of evidence of the effects. A second method applies an additional set of weights to the effects based on the strength of evidence (R-SEED) rating. For this approach, we developed a weighting system that gave more weight to the effects from more rigorous and highly-rated designs. The effects at each level in the R-SEED rating system were given twice the weight of the level below it.

[5] R-SEED originated in work being conducted on a national evaluation funded by a separate department of the federal government. This evaluation also involved coding and analyzing findings from multiple studies of a wide variety of intervention models using a broad spectrum of designs. Project LAUNCH was the first set of data with which the final version of R-SEED was used.

Challenges to the Cross Site Evaluation

The data from the local evaluations presented challenges to the information on outcomes of Project LAUNCH provided by the Cross Site Evaluation. These include:

- The representativeness (or lack thereof) of the full set of LAUNCH service strategies on which outcomes were collected.
- The preponderance of one-group designs in the local evaluations of outcomes of LAUNCH services and systems change activities.
- The quality of the outcome measures used in the evaluations.
- The limited level of funding for local evaluation designs involving comparison groups or administrative datasets.

Data reported by the LAUNCH grantees on their services and systems change activities (Volume I of the Cross-Site evaluation) demonstrate the extent to which the full set of LAUNCH prevention and promotion strategies is being implemented across the grantees. Each grantee had to make strategic decisions about which of the key LAUNCH promotion and prevention strategies and systems change activities and initiatives to focus their efforts and budget on, which resulted in most grantees implementing some but not all of the key LAUNCH promotion and prevention strategies. At the same time, the data suggest that, across the grantees, each of the LAUNCH strategies, as well as systems change activities, is being implemented by a substantial proportion of grantees. If the logic underlying Project LAUNCH posits that outcomes depend on the implementation of these strategies, then the data on implementation support the validity of the argument that the local evaluation should expect to see outcomes for providers, parents, and children.

At the same time that there is evidence of a high level of implementation of the key Project LAUNCH strategies, the local evaluations were only able to collect data on the outcomes of a subset of these strategies. When the available outcome data from the local evaluations are compared with the full set of services and systems work being implemented by the LAUNCH grantees, we find that a small percentage of the LAUNCH activities are represented in the data. This limits the extent to which the results can be generalized to the LAUNCH program overall. Nearly all of the grantees (23 out of 24) reported outcome data on only one or two key LAUNCH strategies, whereas these same grantees were implementing four or five strategies in their communities.

The outcome data from the local evaluations were generated by designs that varied in terms of (a) the overall design (comparison group design, pre-post design, post-only design); (b) whether baseline and post-test data were collected on the same sample and using the same measure; (c) the type of outcome measures used (self-report, rating by an observer, direct assessment), norm-referenced, published, or locally-developed measures; and (d) the sample size and representativeness of the study sample. Nearly all of the local evaluations depended on pre-post or post-only designs for estimating the effects of LAUNCH-supported services. For the first three years of the LAUNCH initiative, SAMHSA provided limited guidelines on preferred designs for the local grantee evaluations. At the end of year 3 (June 2011), SAMHSA introduced more specific guidance

intended to improve the rigor of the local evaluation activities across multiple promotion and prevention strategies.

Finally, most of the local evaluations used non-normed measures of parent and child outcomes. In some cases (e.g., socio-emotional development of children), there are few standardized measures available. In other cases, evaluators either chose to develop their own questionnaires or they selected existing measures that had no or little information available on their psychometric properties.

The limitations of the local evaluations—the fact that local evaluations did not include all of the service and systems change activities being carried out by the grantee, the low-rigor designs, and non-standardized measures—may be related not only to the amount of guidance and technical assistance provided to the grantees on evaluation design from the start of Project LAUNCH but also to the relatively limited budget for local evaluations. Grantees had to make strategic decisions about the services or systems work that their local evaluations should target, since grantees could prioritize no more than 20 percent of their funding for data collection and local evaluation.

Implementation Context for the Cross-Site Evaluation of the Outcomes of Project LAUNCH

The first volume of this report (Gwaltney, Goodson, Pfefferle, and Walker, 2014, in review) presents a profile of how each of the key LAUNCH promotion and prevention strategies has been implemented by the three cohorts of grantees. This profile provides a context for understanding the outcome data for each of the LAUNCH strategies.

Altogether, 26 home visiting programs were supported across the 21 LAUNCH grantees who chose to implement home visiting as one of their service strategies. Consistent with SAMHSA guidelines, the majority of these home visiting programs (65%) were evidence-based models. The remaining were either locally-developed models or public health home visiting programs already operating prior to LAUNCH. The majority of the grantees provided training or other program enhancements intended to integrate mental/behavioral health more fully into the home visiting program. This was achieved through staff training focused on increasing staff knowledge and awareness of issues related to maternal and child mental health and of apppropriate referrals for individuals identified as having mental or behavioral health concerns and through funding mental health consultation for the home visitors. The majority of the LAUNCH grantees (92%) supported at least one family support/family strengthening program. Across all three cohorts, these grantees altogether supported 52 different family support programs. Between 65 percent and 70 percent of these family strengthening programs were newly-initiated in the LAUNCH community. Many of the newly-initiated family support programs focused explicitly on maternal and child mental health. LAUNCH also enhanced both new and existing programs with additional training on issues related to maternal and child mental health and on appropriate assessment of maternal and child mental and behavioral health.

Eighteen (two-thirds) of the 24 grantees supported mental health consultation in child care or preschool settings. Altogether, these 18 grantees supported a total of 20 different early childhood mental health consultation programs. In addition, 10 of the 24 grantees (42%) implemented early childhood mental health consultation in elementary schools. One grantee implemented two different approaches with different schools, resulting in a total of 11 programs.

One of the pillars of the LAUNCH model was a dual focus on systems development as well as service delivery. Among other goals, LAUNCH grantees were expected to help build collaborative relationships among provider organizations across disciplines or systems. To this end, grantees also engaged in activities to enhance the state, tribal, and community early childhood delivery systems and the legislative and organizational policies and practices that influence children's developmental and health outcomes. Grantees implemented six types of systems change activities: partnership development; policy/infrastructure development; data and information systems development; developmental screening/assessment at a population level; workforce development; and public awareness. All grantees were involved in at least one type of systems activity at the state, tribal, and community levels. The majority (83%) of grantees engaged in three or more types of systems activities at the community level. Sixty-one percent (61%) of grantees in Cohorts 1 and 2 implemented three or more types of systems activities at the state level.

Conclusions and Discussion

What Did We Learn from the Cross-Site Evaluation?

Implementation data reported by the LAUNCH projects (see Volume I of this final report—Gwaltney, Goodson, Pfefferle, Walker, 2014) indicate that Project LAUNCH grantees were able to implement mental health enhancements to early childhood services and planning and, in some cases, initiate systems changes. The outcome data available at this point in Project LAUNCH suggest that the providers, parents, and children who are part of the LAUNCH initiative are on positive trajectories, even though the fact that the evaluation studies do not include comparison groups means that LAUNCH cannot claim to be the causal agent of this growth. To have evidence to make statements about LAUNCH effectiveness requires that future evaluations test whether growth in program participants is greater than growth in similar families not in LAUNCH. Having this stronger evidence will not only provide evidence to support the LAUNCH model, it also will allow for additional testing of key questions about implementation practices or systems partnerships that are the most effective for creating the kinds of changes LAUNCH was developed to address.

Further, the implementation and outcome data suggest that it takes time for grantees to implement effective programs. The Cohort 1 grantees found larger effects, relative to the other cohorts. This may indicate that rigorous evaluations would be most appropriate in the second half of the LAUNCH grants, after the grantees have had time to learn about how the programs are being implemented and to potentially institute program improvements, and time to design and implement more rigorous evaluations with appropriate comparison groups. The overall positive direction of the results is mirrored in the results from the SAMHSA Provider Survey and Parent Survey.

Taken in aggregate, the results of the cross-site analyses provide glimpses into possible effects of LAUNCH. Taken at face value, the average effects of LAUNCH strategies are moderate-to-small, with the smallest average effects on children. This report has made clear, however, that (a) nearly all of the results are generated by less rigorous studies; (b) less rigorous studies tend to have larger positive effect sizes; and (c) these study designs leave open the very real possibility that the changes observed in providers, parents, and children are caused by something other than LAUNCH—for example, history (other events happening in the lives of the respondents) or development (especially in the case of child outcomes).

The examination of potential predictors of variation in effects focused on two implementation indices—one related to the depth of the integration of behavioral health in individual LAUNCH programs, and the second related to the breadth of initiatives to improve the child service system at the community, tribal, and state levels. These analyses were generally inconclusive, although there were a few findings that suggested that these indices were related to the size of the observed effects.

Lessons for Project LAUNCH Going Forward

There are potential lessons for Project LAUNCH going forward. First, as would be predicted by the Project LAUNCH logic model, the outcomes for children had the smallest average effect sizes. Whereas improvements for providers and parents may be achievable in the sort-term, improving child outcomes takes longer. The evidence from early childhood education research suggests that interventions that work directly with children are more likely to result in effects on children, compared with the effects of interventions that attempt to affect child outcomes through parents. Since child outcomes are the key goal of Project LAUNCH, the focus of the LAUNCH model could be revisited to determine if greater emphasis should be placed on direct services for children. Second, the relationships shown between the breadth of integration activities in integration of mental health and some outcomes suggest that how grantees approach mental health integration could make a difference. In particular, general staff training on topics related to child socio-emotional development and training on child assessment are not as strongly related to outcomes as are using mental health clinicians to work directly with staff; and, in the case of early childhood and school settings, focusing the mental health consultation on the broader classroom environment as opposed to individual children and families, may be beneficial. Third, the relationship of systems change activity to outcomes suggests that additional attention could be paid to the breadth and depth of grantees' efforts to bring about systems change at the state, tribal, and community levels.

All of these hypothesized lessons about Project LAUNCH need to be validated in another round of outcome analysis that includes more grantees and more years of implementation of the later cohorts of grantees. With additional research, we can, with more confidence, use the findings to validate the effectiveness of Project LAUNCH in improving outcomes for young children and their families and enhancing all levels of the early childhood system.

Table of Contents

1. INTRODUCTION TO PROJECT LAUNCH

1.1 Overview of Project LAUNCH

1.1.1 Program Goal and LAUNCH Promotion and Prevention Strategies

Project LAUNCH (Linking Actions for Unmet Needs in Children's Health) is a federal grant program administered by the Substance Abuse and Mental Health Services Administration (SAMHSA). The goal of this national program is to promote the physical, social, emotional, cognitive, and behavioral health of young children from birth to eight years of age (SAMHSA, 2008). Grantees who receive funding from SAMHSA subsequently provide funding to a pilot community to pursue goals of improving early childhood systems and early childhood services.

The goals of Project LAUNCH are to:

- Increase the integration of mental and behavioral health into early childhood services, including home visiting, family strengthening and parent education, early childhood education and care, and primary care;
- Expand the use of culturally-relevant, evidence-based prevention and wellness promotion practices (EBPs) in home visiting, family strengthening and parent education, early childhood mental health consultation, and integration of behavioral health in primary care;
- Increase access to screening, assessment, and referral to appropriate services for young children and their families in a range of child-serving settings; and
- Improve coordination and collaboration across local, state, tribal and federal agencies serving young children and their families.

Project LAUNCH grantees are guided by Young Child Wellness Councils (YCWCs) that engage key leaders in overseeing Project LAUNCH activities. Grantees form YCWCs at both the state and tribal (Cohorts 1 and 2) and community levels (all cohorts) engage key leaders in developing a strategy and plan for improving outcomes for young children and their families. Grantees could establish new YCWCs or integrate the functions of the YCWCs into existing advisory groups or councils whose focus is young child wellness. At a minimum, YCWCs are expected to have representatives from health, mental health, child welfare, Medicaid, substance abuse prevention, early childhood education and care and state education departments, Title V administering agencies (if applicable), and representation from families in the target population (SAMHSA, 2008; 2009; 2010).

To determine the programs and services they would implement, grantees engaged in a two-step planning process in the early months of the grant. They began by conducting "environmental scans" at the state/tribal and community levels to identify gaps in the existing systems and programs for children, birth to 8 years of age and their families, and perceived community needs. Grantees then used the results of their environmental scans to develop a strategic plan for supporting systems changes and addressing the gaps in services for families and children. The strategic plan identified the evidence-based programs Project LAUNCH grantees would implement within communities.

1.2 Project LAUNCH Grantees

As of Summer 2014, SAMHSA has funded five cohorts of grantees. This outcome report includes data from the first three cohorts, funded in 2008 (6 grantees), 2009 (12 grantees), and 2010 (6 grantees). The grants in the first and second cohorts were awarded to the Title V Maternal and Child Health agency at the state level or a tribal government. Each grantee identified a target community in which to implement evidence-based programs and services for young children and their families. The six grants in the third cohorts were awarded directly to a local government, community agency, or other non-profit entity. Together, the first three cohorts include 24 Project LAUNCH programs operating in 21 states, the District of Columbia, and one tribal nation. See Appendix A for a list of the grantees in the first three cohorts.

While geographically diverse and varied with respect to their target populations, LAUNCH communities were all selected because of their high need for services—e.g., children and families had significant risk factors, services were not sufficient to meet all needs, and the communities had significant health and economic disparities (see Gwaltney, Goodson, & Walker, 2012). For example, families living below the poverty level were 40 percent higher in LAUNCH communities than the country overall (14.4% vs. 9.9% in 2009), and 18.5 percent of all births in LAUNCH communities were to women receiving late or no prenatal care compared to 7.0 percent in the U.S. On average, 25.4 percent of individuals in LAUNCH communities spoke a language other than English at home, compared to 19.6 percent of U.S. residents. The proportion speaking a language other than English ranges from 1.0 to 74.9 percent across all LAUNCH communities.

2. CROSS-SITE EVALUATION OF THE OUTCOMES OF PROJECT LAUNCH

2.1 Objectives of the Cross-Site Evaluation

In addition to awarding grants to states, tribes, or communities to implement the Project LAUNCH model, SAMHSA funded an independent Cross-Site Evaluation of Project LAUNCH, with oversight of the evaluation provided by the Office of Planning, Research and Evaluation (OPRE) within the Administration for Children and Families (ACF) in the U.S. Department of Health and Human Services. Abt Associates Inc. was awarded a five-year contract to conduct the Cross-Site Evaluation of the initiative in consultation with SAMHSA. The evaluation was charged with (a) documenting the implementation of the Project LAUNCH model by the grantees, including supports for programs/services in the community in five key areas and activities to improve the service system; (b) documenting the extent to which Project LAUNCH changed outcomes for providers, parents and children in the LAUNCH communities; (c) documenting the effectiveness of collaboration and coordination between state, territorial, tribal, and locally-based service networks; and (d) providing technical assistance to Project LAUNCH evaluators on the design and implementation of their local, project-specific evaluations.

The first three years of the Cross-Site Evaluation focused on program implementation. Three previous evaluation reports provided cumulative descriptions of the implementation of services and systems change activities by the LAUNCH grantees (Gwaltney, Goodson, Walker 2013; Goodson, Gwaltney, Walker, 2014; Gwaltney, Goodson, Pfefferle, and Walker, 2014). This evaluation report summarizes the results of studies of LAUNCH outcomes, conducted by the local evaluators. The data on outcomes were reported by LAUNCH grantees in annual end-of-year evaluation reports. Conducting grantee-level studies of community and state/tribal efforts was one part of the requirements for award of a Project LAUNCH grant (see Appendix B for language in RFA for Cohorts 1-3 regarding local evaluations). The Cross-Site Evaluation also uses results from a set of separately-funded studies that compared community-wide child outcomes in the LAUNCH community to outcomes in one or more comparison communities. Six grantees from Cohorts 1 and 2 were awarded additional evaluation funds, based on a competitive process. These grantees together implemented ten different studies looking at community-wide outcomes. (See Appendix C for further information on designs of special studies.) Because of reporting schedules for these studies, the current report includes findings from three of the ten studies.

The local evaluation studies measured outcomes for providers, parents, and/or children as well as local and state systems changes. The number of annual evaluation reports, and therefore the potential number of findings, available varies by grantee cohort. Each Cohort 1 grantee had annual reports from five years of implementation, Cohort 2 grantees had reports from four years of implementation, and Cohort 3 grantees had reports on three years of implementation.

2.1.1 Research Questions for the Outcomes Evaluation

The following evaluation questions provide the framework for the Cross-Site Evaluation of the outcomes of Project LAUNCH:

- How has health and well-being improved for young children (birth – 8 years) participating in LAUNCH-supported services?
- How has the local child services system changed during the Project LAUNCH grant implementation?
- How have the state or tribal child services systems changed during the Project LAUNCH grant implementation?

In addition, the Cross-Site Evaluation added two questions about outcomes for providers and parents, although these were not among the original research questions:

- How have knowledge and practices changed for providers of LAUNCH-supported services?
- How have the health and well-being improved for parents of children participating in LAUNCH-supported services?

Another Cross-Site Evaluation question was addressed in the first volume of this report on Project LAUNCH implementation:

- How have child and family services in the community been enhanced?

2.2 Context for the Measurement of Outcomes of Project LAUNCH

The research question about improvements in children's developmental outcomes is the key question for Project LAUNCH. Child health and well-being, especially mental and behavioral health, is the over-riding objective of the program. The logic model for Project LAUNCH assumes that outcomes for children will take longer to appear, since they are primarily being affected through changes in attitudes and practices of their parents and of their child care and health care providers. In addition, providing evidence of the outcomes of site-based, multi-strand initiatives such as LAUNCH, whether the outcomes are for providers, parents, or children, is challenging. Further, estimating child outcomes may be the most challenging—studies of interventions for young children face measurement challenges, for example, the lack of standardized measures in the area of socio-emotional development and the design problem of measuring change at the same time that children are developing.

The earlier reports on implementation make clear that LAUNCH grantees are implementing a broad array of evidence-based programs and national program models in their communities that address the key promotion and prevention strategies and are involved in initiatives to improve the child services system at the community, tribal, and state levels. At the same time, these reports underscore the variation among the grantees—variation in the number of promotion and prevention strategies being implemented and in the approaches grantees are using to enhance services and change local and state systems. Overall, the reports on implementation document sufficient levels of enhancements to services and systems that are hypothesized as the necessary, if

not sufficient, precursors of improved health and well-being for the young children in the LAUNCH community. The documented successes of the LAUNCH grantees in delivering evidence-based services to families and initiating system changes makes it reasonable to ask questions about the possible consequences of these activities for key agents in children's lives—parents and service providers (primary care physicians, child care providers, and other early childhood service providers). That is, have the Project LAUNCH activities changed the knowledge and practices of the adults who care for children, especially as it relates to child mental and behavioral health? Finally, if LAUNCH has changed these key adults, it is reasonable to examine the evidence for what is the critical outcome for Project LAUNCH—improved well-being for young children in the LAUNCH communities.

At the same time, the variation across grantees in the types and intensity of strategies being implemented poses a challenge to cross-site analysis. The cross-site analysis is further complicated by the variation across grantees in the designs that produced the outcome data. Local evaluations varied in terms of (a) which service and systems change strategies were studied; (b) the designs used (comparison group design, pre-post design, post-only design); (c) whether baseline and post-test data were collected on the same sample and using the same measure; (d) the type of outcome measures used (self-report, rating by an observer, direct assessment), norm-referenced, published, or locally-developed measures; and (e) the sample size and representativeness of the study sample. Also, the local evaluations varied in comprehensiveness—the number of individual LAUNCH-supported programs that were assessed and the coverage of the LAUNCH prevention and promotion strategies that were represented in the evaluations.

For the first three years of the LAUNCH initiative, SAMHSA provided general guidelines on preferred designs for the local grantee evaluations. During this time period, the technical assistance staff from the CSE worked with evaluators to try to identify opportunities for more rigorous studies of their service strategies, e.g., discussing possible comparison groups in the community. At the end of year 3, SAMHSA introduced more specific guidance intended to improve the rigor of the local evaluation activities across multiple promotion and prevention strategies (see Exhibit 1). The SAMHSA guidance provided additional incentives for local evaluations to undertake more rigorous evaluation of the service or systems changes being implemented with LAUNCH funding, and the CSE staff provided technical assistance to evaluators on possible study designs and on other aspects of more rigorous studies, such as the use of established instruments to measure outcomes.

Exhibit 1. Summary of SAMHSA Expectations for Grantee-Specific Local Evaluations (2011)[a]

Sample	Expectations for All Grantees	Cohort-Specific Expectations
Provider outcomes	For all programs, assess provider outcomes annually, using SAMHSA Provider Survey[b] or other measure aligned to program objectives Encouraged to implement a quasi-experimental design for at least one key service	Cohort 2 & 3: For 1- 3 key programs, conduct pre-post assessment of provider outcomes using a validated outcome measure. (Cohort 1 grantees encouraged to implement this level of evaluation.)
Parent outcomes	For all programs delivering services directly to parents, assess parent outcomes annually using SAMHSA Parent Survey[c] or other measure aligned to program objectives	Cohort 2 & 3: For 1- 3 key programs, conduct pre-post assessment of provider outcomes using a validated outcome measure. (Cohort 1 grantees encouraged to implement this level of evaluation.)
Child outcomes	For 1 to 3 key programs that deliver services directly to children: measure child outcomes annually pre-post Encouraged to implement a quasi-experimental design for at least one key service	None
Systems outcomes	Annual narrative information about outcomes of systems change initiatives	Cohort 3: valid measure of collaboration (state and local) at baseline and annually thereafter

[a] Appendix D provides the detailed list of evaluation expectations by cohort.

[b] SAMHSA Provider Survey was co-developed by Cohort 1 evaluators, SAMHSA, and the CSE to assess changes in provider knowledge and practices during the time LAUNCH was involved with the program. The SAMHSA Provider Survey is a retrospective pre-post survey that asks providers to indicate the extent of change they have experienced at the time of the survey relative to the time before LAUNCH became involved in the provider's specific program in four areas: knowledge of children's socio-emotional development, knowledge of referral options in their community for children identified as having behavioral or mental health concerns, use of mental health consultation for children with behavioral or mental health concerns, and use of screening in the provider's program. A copy of the SAMHSA Provider Survey is shown in Appendix E.

[c] SAMHSA Parent Survey was co-developed by Cohort 1 evaluators, SAMHSA, and the CSE to assess parent satisfaction with help from LAUNCH program in making them more effective parents and in promoting their child's learning and development. Parents were asked to complete the survey at the end of a program or, for multi-year programs, at the end of each program year. A copy of the SAMHSA Parent Survey is shown in Appendix E.

2.2.1 Methodological Decisions in Designing the Cross-Site Outcomes Evaluation

The design used for the Cross-Site Evaluation is a meta-analysis of the outcomes of individual LAUNCH-supported programs across the 24 LAUNCH grantees in Cohorts 1-3 as of September 2013. This design was selected after consideration of a number of different designs (Exhibit 2):

1. *The level at which outcomes are measured:* One approach to the evaluation would be to focus on questions about the effects of individual programs for providers, parents, and/or children as a result of LAUNCH supports/enhancements to the programs. Alternatively, the evaluation could focus on questions about the overall effect of LAUNCH activities on community-level outcomes for all parents and/or children in the LAUNCH community, not just those participating in specific LAUNCH-supported programs.

2. *The role of the local evaluator vis-à-vis the Cross-Site Evaluation*: The Cross-Site Evaluation could be designed to include its own data collection on program-level or community-level outcomes, or the Cross-Site Evaluation could draw on findings from studies conducted by each of the grantee-specific local evaluations. Grantees might also be required by SAMHSA to use similar local evaluation designs and outcome measures in their evaluations. For example, SAMHSA might have required all grantees to conduct quasi-experimental studies of their family strengthening programs, using the Parent Stress Index as an outcome for parents and the Eyberg Child Behavior Inventory as a child outcome. This would allow the Cross-Site Evaluation to combine and summarize data, with the possibility of pooling individual data across grantees. On the other hand, SAMHSA might have allowed grantees to use different designs and outcomes. In this situation, the Cross-Site Evaluation would take the form of a meta-analysis, with standardized effect sizes calculated for different outcomes to allow for cross-site analyses.

Design options for the Cross-Site Evaluation of Project LAUNCH outcomes were discussed among SAMHSA and ACF staff, as well as with a panel of expert consultants. The designs in Exhibit 2 were considered, and all but the design in Column 1 of the exhibit were ultimately ruled out for different reasons. The option that involved requiring local evaluations to conduct more rigorous evaluations of particular programs was deemed not possible, given the resources that would be required to conduct these evaluations and because specific evaluation requirements (resources, allowable designs, etc.) were not included as part of the Request for Applications for Project LAUNCH grants. The funding of the Special Studies represented an effort to address these barriers for a small number of grantees who were interested in conducting more rigorous community-level studies and who had data available to do so.

The option that involved local evaluators using similar designs and measures was ruled out because of the variety of programs being implemented by LAUNCH grantees, even within the five LAUNCH prevention and promotion strategies. Additionally, some grantees were initiating new programs and others introducing enhancements, and the research questions for these activities would be different. Programs also served different types of families and children, using different theoretical frameworks and different lengths of exposure.

The option that involved the cross-site evaluator designing and conducting evaluations of individual programs addressed at least part of the resource concern but not the concern raised by the extensive variety of program models being implemented. The option that involved the cross-site evaluator conducting community-level studies was considered to hold the most promise in terms of aligning with the LAUNCH program model; however, this design would only be as good as the community-level outcomes that were available. After some discussion and investigation, it was determined that there was not an appropriate child outcome that would be available in extant data across all LAUNCH communities and possible comparison communities. Further, the extant data would need to be available for the LAUNCH-defined community, which may be one or more counties, selected neighborhoods within a larger metropolitan area, or several ZIP codes. There are only a small number of outcomes that the CSE could count on being available for virtually all grantees (except the tribal grantees) and potential comparison communities and that could be

Exhibit 2. Design Options Considered for the Cross-Site Evaluation of Outcomes of Project LAUNCH

Role of Evaluator	Program-level Studies			Community-level Studies		
Local evaluator	Local evaluators choose own designs, measures for programs	Local evaluators agree to use common and internally valid designs, common outcome measures within each program strategy	Act as local facilitator for CSE	Local evaluators design own internally valid community-level evaluations and choose own community outcome(s) to look for different trends over time	Local evaluators agree to implement common internally valid community level design and common outcome measures	None
Cross-site evaluator (CSE)	CSE extracts raw means to calculate – treatment-comparison (T-C) or pre-post differences and calculate standardized effect sizes for meta-analysis of outcomes within strategies	CSE pools individual data across common designs and calculates overall LAUNCH effect within strategies	CSE designs and conducts program-level studies in each community, using same design and outcomes	CSE extracts evaluator-calculated effect sizes for meta-analysis	CSE conducts analysis pooling LAUNCH sites and comparison sites for overall LAUNCH effect	CSE designs and conducts community-level studies across all grantees, using common outcome measure, conducts pooled analysis for overall LAUNCH effect

disaggregated to the community level (i.e., by ZIP code)—i.e., third grade scores on state math and reading tests and child maltreatment. However, neither of these outcomes is well-aligned with the LAUNCH emphasis on children's socio-emotional development. Additionally, third grade represents the oldest end of the LAUNCH focal age range (birth – 8 years), and there was concern, first, that many LAUNCH-supported programs focused on children in their earlier years and, second, that the effects of LAUNCH would be manifested at the end of the five years of implementation.

The cross-site team also explored with SAMHSA and ACF using the RE-AIM framework for the outcomes evaluation. This framework has five dimensions: Reach, Effectiveness, Adoption, Implementation, and Maintenance. While most of the local evaluators showed interest in the RE-AIM framework, they also said it would take time to operationalize measures for each dimension, and it would be particularly challenging to define the "Reach" and "Effectiveness" dimensions in the context of Project LAUNCH. For example, "Reach" requires using a denominator representing the size of the eligible population for each service, and this number is unknown, especially for programs

that target at-risk families. Additionally, the RE-AIM framework was thought to be less applicable to programs enhanced by Project LAUNCH (versus programs that were newly-initiated), and program enhancements were a primary strategy used by LAUNCH grantees. For these reasons, it was deemed infeasible to use RE-AIM for the Cross-Site Evaluation of Project LAUNCH.

These deliberations about possible designs led to the conclusion that the current meta-analytic approach had the most promise as a feasible and informative evaluation design.

2.2.2 Methodology

Given the design decisions made and the available data from the local grantee evaluations, the analysis of the LAUNCH outcome data takes the form of a meta-analysis that combines data from different designs, different samples (providers, parents, children), and different outcome measures. The data available for the meta-analysis reflects outcomes for grantees at different developmental stages: the six grantees in Cohort 1 had completed their five-year grant period; grantees in Cohort 2 were in their fourth year; and grantees in Cohort 5 had completed three years of their grant. Therefore, 18 of the 24 Project LAUNCH sites (75%) had only partially implemented their grants when the meta-analysis was conducted.

Because this is a meta-analysis, it is necessary to put the findings from these different sources into effects on a common metric. Typically, this is achieved by calculating a standardized effect size for each treatment-control (T-C) difference. (The field uses the term "contrast" to refer to a difference on an outcome between a treatment and control group or between two scores for a single group in a pre-post design.) In addition to transforming each contrast into a standardized effect size, we evaluated the strength of the evidence that each effect represents, i.e., the level of confidence we can have that the effect represents a true measure of the effect of the intervention being studied.

Below, we describe our methodology for calculating standardized effect sizes and assigning strength of evidence ratings. We then describe the process by which we combined the outcome data from all of the LAUNCH evaluations and analyzed the results.

2.2.2.1 Extracting and coding data from local grantee evaluations

Obtaining the data for the meta-analysis of LAUNCH outcomes was a two-step process. First, a team of trained reviewers examined all of the annual end-of-year evaluation reports for each of the 24 grantees in Cohorts 1, 2, and 3 and the available reports from the Special Studies. Both steps involved a small team of reviewers who were trained on the data extraction system and on the data coding system. In the first step, two trained reviewers[6] examined each evaluation report and independently identified where outcome data were reported on LAUNCH-supported programs or systems change and where the same outcomes were measured and linked across years. The pair of reviewers compared the outcome data that were identified and came to agreement on the data that would be extracted for the meta-analysis. In the second step, the outcome data were

[6] The eight trained reviewers were paired with different partners across the reviews of the data from the 24 different grantees. Two of the most senior reviewers also served as "reconcilers" when the reviewers did not agree on their reviews or coding of the data.

systematically coded for analysis, using an Excel data entry system that included information on the LAUNCH-supported program on which the outcomes were measured, the characteristics of the outcome measures, the design of the outcome study, the analysis model, and the findings. The two members of the reviewer team coded the data independently and compared their final coding sheets. Where there were differences for which the reviewers could not arrive at an agreed-upon coding, a memo was sent to the senior reconciler explaining the disagreement, and the reconciler determined how the data should be coded. The last step in the coding was for one of the reviewers to create a master review, which was subsequently used for the meta-analysis.

During the coding of the outcome data, it was frequently determined that the design for the outcome study did not generate a "contrast" or difference between two groups. For example, in some evaluations, outcomes were reported at the end of a program, with no baseline or pre-test data on the same respondents or data on an external comparison group. These outcomes were not included in the meta-analysis, since there was no basis for calculating a difference or an effect. Outcome data that could not produce an effect were not coded and not included in the meta-analysis. Most of the Project LAUNCH evaluations included a set of outcomes for which no contrast could be created, because they used two outcome measures that were created in the first year of the LAUNCH initiative to provide common measures for local evaluators to use in the absence of other approaches to measurement. These measures—a Provider Survey and a Parent Survey—were designed as retrospective pre-post measures, where respondents were asked about how much change had occurred in relevant knowledge or behaviors against a baseline. Although the data can be expressed in terms of amount of change, without actual baseline measures, effect sizes for pre-post differences cannot be calculated. Although the data from these surveys could not be included in the meta-analysis, separate descriptive analysis was conducted on the findings from these surveys that were reported by local evaluators.

2.2.2.2 Calculation of standardized effect sizes for analysis

Once the raw data from the evaluation reports were coded into spreadsheets, standard statistical software[7] was used to transform each effect or contrast into a standardized effect size (Hedges' *g*). The resulting effect size represents the size of a treatment-comparison difference relative to the standard deviation of the outcome measure. In the education field, thresholds have been suggested for small (0.15 standard deviations (SD)), medium (0.45 SD) and large effect sizes (0.90 SD) (Lipsey, 1990). More recently, methodologists have stressed that simply applying these conventions concerning the magnitude of observed effects is not sufficient. Instead, the magnitude of effects should be interpreted in the context of other research on similar interventions and in terms of their practical value (Durlak, 2009; Durlak, Weissburg, et al. 2011; Hill, Bloom, Black & Lipsey, 2007). The size of the effects obtained from the LAUNCH outcome analyses were interpreted, where possible, in light of meta-analyses of relevant interventions and other multi-study summaries of findings (see later discussion in Section 5.1.5).

[7] The software program used for the LAUNCH analysis to calculate standardized effect sizes was Comprehensive Meta-Analysis Version 2®.

Accounting for Sample Size. When using standardized effect sizes in the outcome analyses, we weighted each effect size by the size of the sample.[8] The assumption is that the effects that were measured on larger samples were more accurate representations of the true effects. The result is that the effects measured with larger samples have more weight in the analysis.

2.2.2.3 Strength of evidence ratings

Each contrast, in addition to being transformed into a standardized effect size, is assigned a rating by the Cross-Site Evaluation team that communicates the level of confidence that the difference can be attributed to the intervention being tested, i.e., that the difference represents a valid measure of a program effect. This rating is labeled as the "strength of evidence" of the contrast, which represents the internal validity of the estimate of the effect. The most rigorous designs, such as randomized control trials, provide the strongest evidence of a treatment effect, because the randomization ensures that the only statistical difference between the two groups being compared is the treatment. Contrasts from studies using other designs, such as quasi-experimental designs or pre-post designs, receive lower strength of evidence ratings because of potential competing hypotheses that could explain the difference, such as time or development. Knowing the strength of evidence for contrasts in a meta-analysis such as this one for Project LAUNCH allows us to describe the extent to which the result of the meta-analysis can be seen as causal evidence of the effects of LAUNCH.

The system used to assign a rating of the strength of the evidence was the Rating System for Strong and Emerging Evidence Designs (R-SEED)® (Goodson, Price, Darrow et al., in development). R-SEED was selected[9] because it encompasses the full range of designs, from the most rigorous (randomized control trials) to designs that do not involve external comparison groups (e.g., pre-post studies). Although there are other evidence rating systems, such as the those used by the What Works Clearinghouse (WWC) from the U.S. Department of Education, the Clearinghouse for Labor Evaluation and Research (CLEAR) from the U.S. Department of Labor, and the United States Preventive Services Task Force, these systems focus on evaluating evidence from more rigorous designs. Because LAUNCH grantees were given broad latitude in designing evaluations, had limited resources to conduct more rigorous designs, and were unable in many cases to identify a matched comparison group, most of the evidence generated by the LAUNCH evaluations is from less rigorous studies. Therefore, it was important to use an evidence rating system that could distinguish not only very rigorous designs from those that are less rigorous, but could also differentiate evidence produced by a variety of less rigorous designs.

As described above, the data from the Project LAUNCH evaluation reports were systematically coded into a spreadsheet that included nearly all of the information needed to calculate an R-SEED

---

[8] The weighted effect size was calculated by multiplying the effect size by the inverse of the effect size standard error squared.

[9] R-SEED originated in work being conducted on a national evaluation funded by a separate department of the federal government. This evaluation also involved coding and analyzing findings from multiple studies of a wide variety of intervention models using a broad spectrum of designs. Project LAUNCH was the first set of data with which the final version of R-SEED was used.

rating (see Box B); as noted above, a standardized effect size was calculated separately using standard software. In the same way, an R-SEED rating for each contrast was generated by a program that weighed the various design features on the coding sheets. Six R-SEED ratings could be assigned to contrasts in the LAUNCH data:

- Strong evidence,
- Emerging evidence,
- Limited evidence,
- Limited evidence with reservations,
- Weak evidence, and
- Does not meet R-SEED evidence standards[10].

The R-SEED rating for different designs are described briefly below; fuller descriptions of the R-SEED standards are provided in Appendix F.

Box B
Summary of R-SEED® Standards for Strength of Evidence
(full standards provided in Appendix F)

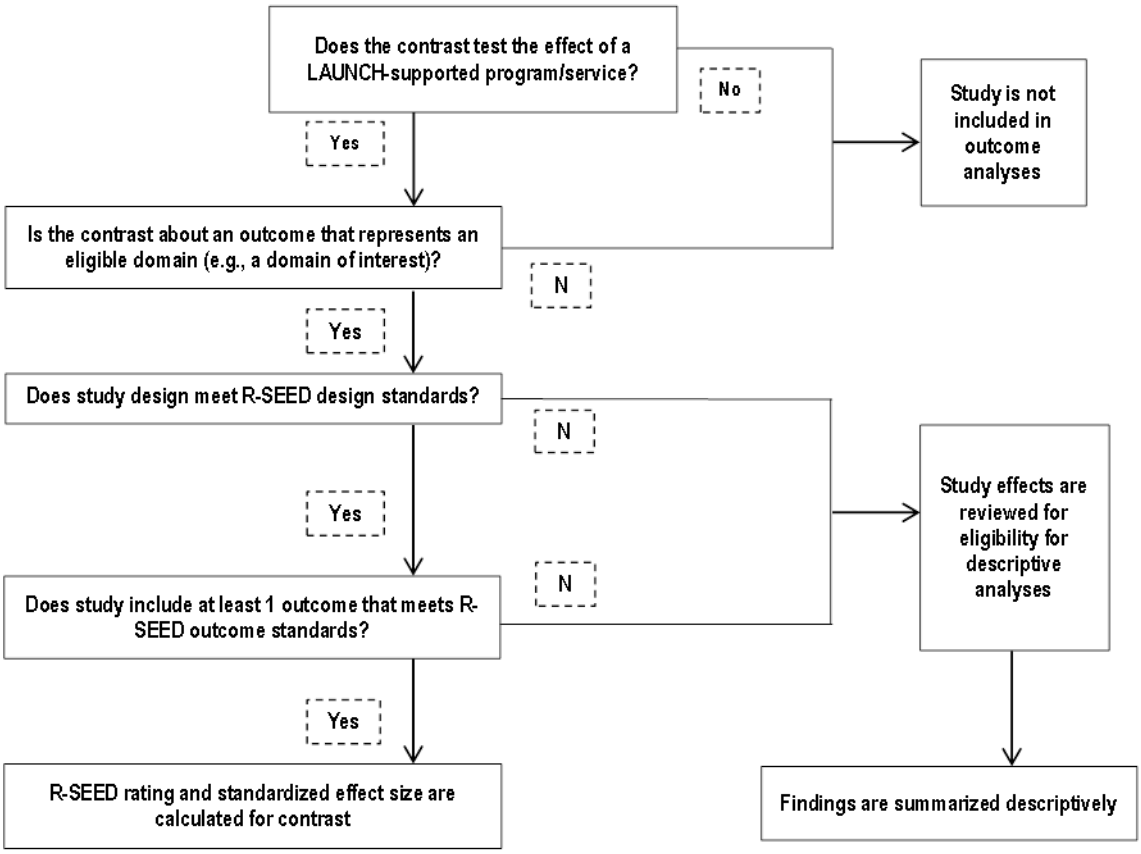
Strong evidence
• Well-design *randomized study* with low attrition • Well-designed *quasi-experimental study* with baseline equivalence of the T and C analysis sample • At least one outcome that meets all R-SEED standards for reliability and face validity
Emerging evidence
• *Quasi-experimental study* that does not demonstrate baseline equivalence of the T and C analysis sample but pre & post are highly correlated and analysis controls for baseline • *Interrupted time study* (≥ 3 baseline time points) with same outcome at all pre & post time points, and baseline trend modeled and accounted for in analysis • At least one outcome that meets most R-SEED standards for reliability and face validity
Limited evidence
• *Quasi-experimental study* that does not demonstrate baseline equivalence of the T and C analysis sample but pre & post are moderately correlated and analysis controls for baseline • *Interrupted time study* (≥ 3 baseline time points) that accounts for baseline trend in analysis • *Pre-post study (< 2 baseline time points) with same measure at pre- and post, using normed measure that demonstrates baseline model* • *At least one outcome that meets some R-SEED standards for reliability and face validity*
Limited evidence with reservations (adaptation for LAUNCH R-SEED coding)
• *Same as "Limited" except that Pre-Post studies do not use normed measure*
Weak evidence
• *Pre-post study* (≤ 2 baseline time points) with different measure at pre- and post, z-scored • At least one outcome that meets one of the R-SEED standards for reliability and face validity

[10] In the LAUNCH Cross-Site Evaluation, the R-SEED rating system was adapted by adding a rating of "Limited evidence with reservations." This rating allowed an additional distinction between studies that used outcomes with differing evidence of psychometric adequacy.

Once the R-SEED ratings were generated, the meta-analysis used (a) information on the LAUNCH program that was coded by the reviewers, (b) the standardized effect sizes calculated from the findings coded by the reviewers, and (c) the R-SEED rating generated by the computer program. The meta-analyses used only the effects that met R-SEED evidence standards, although some of the analyses focused on the effects with the higher R-SEED ratings (see discussion of Analysis Models below).

Figure 1 shows the stages in the review process. The process of coding effects for the outcome analysis is as follows. For each contrast reported, information was coded about the intervention being assessed, the outcome measure that was used for that contrast and its characteristics, the sample, the study design, and the findings. Each contrast was then assigned an R-SEED rating. Finally, a standardized effect size was calculated for each contrast.

Figure 1. Coding Procedure for the LAUNCH Outcome Analysis



2.2.2.4 Outcome domains and types of outcome measures

The LAUNCH local evaluations encompassed a variety of measures as part of their "outcome" reports. For example, many evaluators reported on satisfaction—how positive providers and parents were about the services they received or about provider-parent or provider-provider relationships. These measures were not included in the CSE meta-analysis. For providers, parents, and children, local evaluations also reported on a variety of outcome domains that were considered relevant to the programs being evaluated. As part of the data coding, the relevance of the outcome domain was assessed. The following outcome domains were considered relevant for LAUNCH:

- Provider outcome domains:
    - Provider attitudes about child/family mental health (role of provider in assessment, referral; confidence in own knowledge of mental and behavioral health; value of mental health consultants);
    - Provider knowledge (child development, child socio-emotional development and commonly exhibited behavioral concerns; resources in the community for referrals); and
    - Provider behavior (use of developmental assessments; use of mental health consultation)
- Parent outcome domains
    - Parent attitudes (confidence in their own parenting skills, ratings of the difficulties/stresses of parenting, about their own mental health, about family relationships);
    - Parent knowledge (about positive parenting techniques, about their child's education and how to support it); and
    - Parent behavior (discipline approaches, home supports for child's education and learning, family communication patterns),
- Child outcome domains
    - Child cognitive/language development;
    - Child socio-emotional development/behavior (positive social behaviors, negative social behaviors);
    - Child kindergarten school readiness; and
    - Child school achievement.

Types of outcome measures. The local evaluations used a variety of outcome measures that varied in the extent to which they had supporting information on psychometric characteristics. The characteristics of the outcome measure for an effect are a factor in the R-SEED rating of strength of evidence of that effect. Normed measures are considered to be the strongest,

because they not only have adequate reliability, they also provide a metric by which to assess whether positive change in a sample is larger than the change that would be expected on the basis of time alone.[11] Published measures that are not normed typically have information available on the reliability of the measure based on previous administrations of the measure, but do not have population norms. The proportions of contrasts based on standardized/normed, published or locally-developed measures are shown in Exhibit 3. The names of the normed and published measures are provided in Exhibit 4.

- **Provider effects:** Across all strategies, three-quarters of the provider effects were based on published measures. Three effects were based on normed measures.
- **Parent effects**: Not unexpectedly, provider effects were least likely to be measured with standardized measures, since there are few standardized measures of provider knowledge, attitudes, and practices. Across all strategies, about a quarter of the parent effects were based on normed measures, and half were based on published measures. The remaining effects were based on study-developed measures of unknown reliability.
- **Child effects**: Child effects were most likely to be measured using standardized, normed measures (Exhibit 3). Across all strategies, over half of the child effects were based on standardized measures, and the remaining effects were based on published measures. Virtually no locally-developed child outcome measures were used.
- **Systems changes:** Although the Cross-Site Evaluation pointed evaluators to published measures of inter-agency collaboration, published measures of other types of system change outcomes were not identified by evaluators. Half of the systems change effects were based on published measures, and these were all measures of provider collaboration. The other half of the measures was locally-developed measures of systems change (Exhibit 3). The names of the published measures are shown in Exhibit 4.

[11] The usefulness of norms depends on whether the norming sample is relevant to the study sample. Most of the normed measures of child and parent outcomes are normed on what is described as "a nationally representative sample" although some measures also include norms for special subgroups (that were oversampled in the national norming group), such as parents or children receiving services for identified problems. Even when the study sample is more narrow than the national norms, the national norm provides a common metric for comparing outcomes of different populations against the national population.

Exhibit 3. Number of Effects for Three Types of Outcome Measures by Strategy and Sample

LAUNCH Prevention and Promotion Strategy	Provider Outcomes			Parent Outcomes			Child Outcomes		
	Normed measures	Published measures	Locally-developed measures	Normed measures	Published measures	Locally-developed measures	Normed measures	Published measures	Locally-developed measures
Home visiting	0	7	0	0	97	26	16	31	1
Family strengthening	0	0	16	58	50	45	9	3	1
Early childhood mental health consultation: Preschool	3	151	15	0	0	0	0	54	1
Early childhood mental health consultation: School	0	0	0	0	0	0	64	0	0
Integration of behavioral health in primary care	0	10	4	12	10	0	30	8	0
Systems change							0	191	192

Source: Annual end-of-year local evaluation reports 2009 – 2013; Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

Exhibit 4. Standardized/Normed and Published Measures Used for Outcome Studies of LAUNCH-Supported Programs

Provider Outcome Measures	Parent Outcome Measures	Child Outcome Measures
Home Visiting		
• *Teacher Opinion Survey* (Geller, 1998) • *Mental Health Survey* (Green, Everhard, Gordon, Garcia-Gettman, 2006)	• *Parenting Skills Ladder* (Pratt, McGuigan, Katzev, 2000) • *Life Skills Progression* (Wollesen & Peifer, 2006)	• *Ages and Stages Questionnaire (ASQ)* ★ • *Ages and Stages Questionnaire-Socio-Emotional (ASQ-SE)* ★ • *Eyberg Child Behavior Inventory* (Eyberg, 1990) ★
Family Strengthening		
No effects reported	• *Center for Epidemiological Studies Depression Scale (CESD)* (Radloff, 1977; Eaton et al.,2004) ★ • *Brief Symptoms Inventory* (Derogatis, 1993) • *Parent Stress Index (PSI)* (Loyd & Abidin, 1985) • *Parenting Practices Inventory* (Webster-Stratton, Reid & Hammond, 2001) • *Protective Factors Survey (*Counts, Buffington, Chang-Rios, Rasmussen & Preacher, 2010) • *Parent Health Questionnaire* • *Adult Adolescent Parenting Inventory-Version 2 (AAPI 2)* (Bavolek & Keene,1999) ★ • *Posttraumatic Stress Checklist* (Weathers, Litz, Herman, Huska, Keane, 1993) • *Multidimensional Survey of Perceived Social Support* (Zimet, Dahlem, Zimet & Farley, 1988)	• *Ages and Stages Questionnaire—Socio-Emotional (SQ-SE)*[2] ★ • *Eyberg Child Behavior Inventory* (Eyberg, S., 1990) ★ • *Devereux Early Childhood Assessment (DECA) (LeBuffe & Naglieri, 2009)* ★
Early Childhood Mental Health Consultation: Preschool		
• *Teacher Opinion Survey* (Geller, 1998) • *Mental Health Opinions Survey* • *Mental Health Services Survey* (Green, Everhard, Gordon, Garcia-Gettman, 2006) • *Preschool Mental Health Climate Scale* (Gilliam, 2008) • *Preschool-wide Evaluation Tool (PreSET)* (Steed & Pomerleau, 2012) • *Teaching Pyramid Observation Tool (TPOT)*(Hemmeter, Fox, Snyder, 2008) • *Pyramid Infant-Toddler Observation Scale (TPITOS)* • *Protective Factors Survey (*Counts, Buffington, Chang-Rios, Rasmussen & Preacher, 2010)	*No effects reported*	• *Ages and Stages Questionnaire*[2] ★ • *Devereux Early Childhood Assessment (DECA)* (LeBuffe & Naglieri, 2009) ★ • *Dynamic Indicators of Basic Early Literacy Skills (DIBELS)* (Good & Kaminski,1996) • *Elementary Reading Attitude Survey* (McKenna & Kear, 1990) • *Eyberg Child Behavior Inventory* (Eyberg, S., 1990) ★ • *Kansas Early Learning Inventory* (KS State Dept. of Ed., 2001, 2012) • *New Mexico Pre-K Early Learning Outcomes: Essential Indicators (NM State Dept. of Children, Youth & Families)*

Provider Outcome Measures	Parent Outcome Measures	Child Outcome Measures
		• *Ohio Youth Problems, Functioning, and Satisfaction Scales-Version 3.2 (Ohio Scales)* (Ogles,2007) • *Peabody Picture Vocabulary Test (PPVT)* (Dunn & Dunn, 1997) ★ • *Social Skills Improvement Scales* (Gresham & Eliot, 2008) ★ • *Trauma Symptom Checklist for Young Children* (Briere & Runtz, 1989) • *Woodcock Johnson Tests of Cognitive Abilities* (Woodcock, McGrew, & Mather, 2001, 2007) ★
Early childhood mental health consultation: Kindergarten – Grade 3		
No effects reported	*No effects reported*	• *Social Skills Improvement System Rating Scales* (Gresham & Elliot, 2008) ★ • *Dynamic Indicators of Basic Early Literacy Skills (DIBELS)* (Good & Kaminski, 1997) • *Elementary Reading Attitude Survey* (McKenna & Kear, 1990) • *Strengths and Difficulties Questionnaire* (Goodman, 1997)
Integration of behavioral health in primary care		
No effects reported	• *Parent Stress Index (PSI)* (Loyd & Abidin, 1985) • *Center for Epidemiological Studies Depression Scale (CESD)* (Radloff, 1977; Eaton et al., 2004) • *Parent Experiences Survey (Triple-P Positive Parenting Program)*	• *Behavior Assessment System for Children* (Reynolds & Kamphaus, 1993; 2004) ★
Systems		
The Wilder Collaboration Factors Inventory (Derose, P.K., Beatty, A., Jackson, C.A.,2003)		
PARTNER Tool Network Survey (Varda, D.M., Chandra, A., Stern, S.A., Lurie, N., 2008)		
Levels of Collaboration Scale (Frey, B.B., Lohmeier, J.H., Lee, S.W., & Tollefson, N., 2006)		

★ = normed measure

Source: Annual end-of-year local evaluation reports 2009 – 2013; Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

2.3 Analysis Models

As specified earlier, the Cross-Site Evaluation encompasses outcomes for providers, parents, and children, and outcomes for the local and state child and family service systems (see evaluation questions in Section 2.1.1). In addition to providing answers to these questions based on the overall average effect sizes both within strategies (home visiting, family strengthening, mental health consultation in early childhood education and care, and integration of behavioral health in primary care) and for all strategies combined, the analyses explored a small set of hypotheses about implementation factors that might explain variation in the outcomes:

1. Within strategies, are there implementation factors that are associated with variation in outcomes for providers? For parents or children?

2. At the grantee-level, are there implementation factors that are associated with variation in the average effects of different LAUNCH projects for providers? For parents or children?

The average effect sizes for these analyses were calculated using two different methods. The first method calculated average effect sizes using all of the relevant contrasts weighted by sample size, regardless of their R-SEED rating for strength of evidence. We also conducted analyses that took into account the strength of evidence of the effects. One approach we considered was to include in the analyses only the effects assigned higher ratings for strength of evidence (e.g., Emerging or Strong evidence). The usefulness of this approach depended on the number of LAUNCH effects at these levels of strength of evidence. If there were very few effects at these levels, the cross-site analysis would be substantially weakened.

A second method, which could include all contrasts, regardless of their R-SEED rating, would apply a second set of weights to the effects based on the strength of evidence (R-SEED) rating. For this approach, we developed a weighting system that gave more weight to the effects from more rigorous and highly-rated designs. The effects at each level in the R-SEED rating system were given twice the weight of the level below it. For example, contrasts assigned an R-SEED rating of Limited with Reservations were given twice the weight of contrasts assigned a rating of Weak evidence, and so on. This weighting system was guided by our own conceptualization of the shape of the relationship between causal strength and type of design. We did not have the advantage of either prior research that used data weighted by strength of evidence or strong theory about causal claims. In the outcome analyses, individual effect sizes were always weighted by their precision, which was essentially their sample size; effects calculated based on larger samples had greater weight than effects calculated from smaller samples. The key effects analyses also applied weights based on strength of evidence (R-SEED) ratings. We note that this system of weighting by strength of evidence is a methodology that has not been employed in prior meta-analyses, so the analyses using these weights have to be seen as more exploratory compared with the analyses weighting only by the sample size of the effect.

2.3.1 Framing the Cross-Site Evaluation of LAUNCH Outcomes Evaluation

There are three conditions that affect the likelihood that the CSE meta-analysis could be expected to generate findings about the effectiveness of Project LAUNCH:

1. The LAUNCH model has been implemented at the level assumed necessary to lead to positive outcomes for providers, parents, children, and systems. Each of the programs supported by LAUNCH has its own theories about what changes are expected to occur in providers, parents, and/or children as a result of the program services; however, these hypothesized changes are linked to adequate implementation of the services (e.g., adequate amount and quality of training of providers, adequate amount and quality of improved provider interactions with parents or children). If the local or Cross-Site Evaluation data suggest that the program services were being implemented at an inadequate level or intensity, then the services could not be expected to result in changes in the targets of the services.
2. The results from the Cross-Site Evaluation will be generalizable to Project LAUNCH at an initiative level if the local grantees implement programs representing all of the key LAUNCH promotion and prevention strategies and if the local grantee-specific evaluations provide evidence on the full set of programs being supported by LAUNCH. If grantees implement some of the LAUNCH strategies and/or if the local evaluations report data on some of the LAUNCH strategies, then the cross-site findings cannot be assumed to be a valid assessment of the overall LAUNCH approach.
3. Inferences about the extent to which the effects can be interpreted to LAUNCH will depend on the strength of evidence of the effects, which is a measure of the internal validity of the evaluations. The findings will inform us about LAUNCH to the extent that the effects are generated by designs of sufficient rigor.

The next two chapters provide evidence of the extent to which each of these three conditions has been met. Chapter 3 documents the extent to which the LAUNCH grantees have implemented the key LAUNCH promotion and prevention strategies and systems change activities and whether there is support for the validity of the argument that we can expect to see LAUNCH outcomes based on how grantees have implemented services and systems work in their communities. Chapter 4 documents the evidence on the remaining two conditions. First, it presents evidence for whether the local grantee evaluations cover the full set of grantee activities. Second, it discusses the strength of evidence ratings represented in the evaluation findings and the proportion of effects that can be considered sufficiently rigorous or have enough internal validity to support attributions of causality of effects to the LAUNCH program.

3. IMPLEMENTATION CONTEXT FOR OUTCOME EVALUATION

In this section, we provide a brief overview of how each of the key LAUNCH prevention and promotion strategies has been implemented by the three cohorts of grantees (a more comprehensive and in-depth discussion of Project LAUNCH implementation can be found in Volume I of this report—see Gwaltney, Goodson, Pfefferle, and Walker, 2014). This information, which is based on data reported by grantees to the Cross-Site Evaluation's Web-based data system (Web portal) (see Exhibit 5), will provide a context for understanding the outcome data for each of the Project LAUNCH strategies. In this discussion, systems change activities are treated as a LAUNCH strategy for the purposes of summarizing implementation to date.

Exhibit 5. Data Sources for Implementation Context

Cohort (Year initially funded)	Implementation Year 1	Implementation Year 2	Implementation Year 3	Implementation Year 4	Implementation Year 5
Cohort 1 (2008)	Web portal: 9 reporting periods (Fall 2009, Spring 2010, Fall 2010, Spring 2011, Fall 2011, Spring 2012, Fall 2012, Spring 2013, Fall 2013)				
Cohort 2 (2009)	Web portal: 7 reporting periods (Fall 2010, Spring 2011, Fall 2011, Spring 2012, Fall 2012, Spring 2013, Fall 2013)				
Cohort 3 (2010)	Web portal: 5 reporting periods (Fall 2011, Spring 2012, Fall 2012, Spring 2013, Fall 2013)				

3.1 Home Visiting

Twenty-one of the 24 grantees in the first three cohorts worked with at least one home visiting program in their community to elevate staff understanding and awareness of maternal and child mental health. In total, LAUNCH supported 26 home visiting programs. Four of the grantees supported two home visiting programs, and one grantee supported implementation of the same home visiting model in three different organizations. For the majority of these programs (75%), the role of LAUNCH was to enhance an existing program by implementing activities to integrate mental/behavioral health into the program. (The other 25% of the programs were new programs initiated by LAUNCH, most of which had a focus on mental and behavioral health or children and mothers.) Consistent with SAMHSA guidelines, the majority of these home visiting programs (65%) were evidence-based models. The other models included locally-developed models that are in the process of obtaining evidence for being rated as evidence-based, models developed by LAUNCH for specific population subgroups in the community (e.g., fathers), and public health home visiting programs, such as one-time newborn home visits conducted by public health nurses.

Grantees used multiple approaches to bring a greater awareness of mental and behavioral health into the home visiting programs. In all but one of the LAUNCH-supported home visiting programs, LAUNCH provided staff training on topics such as children's cognitive, socio-emotional, and physical

development, treatment options and appropriate referrals for children with behavioral and mental health concerns, engaging parents in programs, and involving parents in their child's growth and learning. In most programs, LAUNCH trained staff on administering and interpreting screeners for child mental and behavioral health concerns. Finally, in half of the programs, LAUNCH grantees funded a part-time mental health consultant to work with the staff in the home visiting programs. The consultants played different roles across the LAUNCH programs—all provided some clinical supervision but some provided reflective supervision to afford home visitors working with highly at-risk families the opportunity to reflect on, sort out, and cope with strong feelings brought on by the demands of their work. The mental health consultants also provided consultation to home visitors on individual mothers or children in their caseloads. In a small number of these programs, the mental health consultant was "embedded" in the program, i.e., the professional consultant and the home visitors were co-located within the same agency, which allowed them to provide more consistent and ongoing supervision within the program.

3.2 Family Strengthening/Parent Education

The majority of the LAUNCH grantees (92%) supported at least one family support/family strengthening program. Across all three cohorts, in Fall 2013—the end of the fifth year of SAMHSA's LAUNCH initiative,[12] these grantees together supported 52 different family support programs. Between 65 percent and 70 percent of the family strengthening programs being supported by LAUNCH were newly-initiated in the LAUNCH community.

These family strengthening programs varied in their focus. The majority of the LAUNCH-supported programs (76%) involved offering parent groups to the entire community, without specific eligibility criteria. These programs included evidence-based program models such as Incredible Years, Strengthening Multiethnic Families and Communities, Triple-P and Centering Pregnancy/Parenting, as well as locally-developed approaches. Other family strengthening programs worked with families with identified concerns about a child's mental or behavioral health. Most of these programs use national models, such as Parent Child Interaction Therapy, Trauma Recovery and Empowerment Model, Trauma-based Cognitive Behavior Therapy, and Primary Project. A third category of family strengthening activities involved assessing family needs and providing referrals to appropriate services. Three of the LAUNCH programs fell into this category.

Within the family strengthening strategy, LAUNCH grantees had three roles. First, grantees introduced new programs into the community whose content focused on maternal and child mental health. Second, LAUNCH enhanced both new and existing programs with additional training on issues related to child socio-emotional development. Third, grantees provided training and support for assessment of maternal and child mental and behavioral health as part of the programs. Across all cohorts, around a quarter of the LAUNCH-supported family support programs conducted child screening and fewer programs conducted parent screening. (The fact that screening was part of

[12] Grantees were at different stages of implementation in 2013: grantees in Cohort 1 were completing their last grant year; Cohort 2 grantees were ending their fourth year; and grantees in Cohort 3 were completing their third year.

fewer family support programs, compared with home visiting programs, could be the result of differences in how these two types of programs were implemented. Home visiting programs typically include longer-term, direct work with children and parents, while family support programs more often involve only working with parents and often for a short and limited time period (e.g., a set number of sessions). Programs reported that screening did not seem as appropriate in the context of the shorter-term, less intensive relationships with families in family support programs. Among the different types of family support programs, screening was more likely to be part of navigation and family coordination programs.

3.3 Early Childhood Mental Health Consultation: Preschool Settings

Eighteen (two-thirds) of the 24 grantees implemented early childhood mental health consultation (ECMHC) in child care or preschool settings. Two grantees implemented ECHMC using two different models, resulting in a total of 20 LAUNCH ECHMC programs. All of the programs included providing program staff with general consultation from clinically-trained mental health professionals on topics such as typical and atypical child development, social-emotional development, or classroom environments that promote the development of social and emotional skills. Consultants also provided consultation on programmatic issues such as staff relationships, parent-staff relationships, or programming and curriculum, which typically included observations of the classroom.

Consultants also provided more targeted, child-specific consultation with staff about individual children of families identified by the program staff as having mental or behavioral health concerns. The consultant might observe or assess the child, might meet with the child's parents as well as teachers, suggest strategies or make recommendations to address issues (either specific to the child or to support a positive learning environment that promotes healthy social-emotional development) and could make referrals for additional evaluation or services. Finally, the most intensive consultation activity, offered by a small number of the consultation programs, was short-term mental health treatment for a child or a parent-child pair.

In half of the programs, the mental health consultation was guided by a framework or model. Three programs reported using the Georgetown Model in their mental health consultation, three reported using the Pyramid Model from the Center on the Social and Emotional Foundations for Early Learning (CSEFEL) (Fox, Dunlap, Hemmeter, Joseph & Strain, 2010), two programs used Positive Behavior Intervention and Supports (PBIS), and one program used Triple P. One other program used a locally-developed consultation model based on the Classroom Assessment Scoring System (CLASS), an observation measure of the quality of classroom practices. The rest did not report having a named model. Seven programs supported implementation of a new socio-emotional early childhood curriculum in the classroom, including the Incredible Years (three programs), and Second Step and the Devereux Early Childhood Assessment Preschool Program (one program each).

The mental health consultants were clinically-trained and had specializations in early childhood mental health and development. All of the consultation programs offered the services of the mental health consultant on an as-needed basis. In three of the programs, this was the way that the

consultant was available to the early childhood programs. In eight of the programs, the consultant was physically present at an early childhood site one day a week. In three programs, the mental health consultant was sited at the early childhood program on a more full-time basis. Across the grantees implementing mental health consultation, a quarter worked with programs from two auspices—child care and either Head Start or school district preschool programs. Among the other programs, two-thirds worked with child care programs, a third worked with Head Start programs, and 10 percent worked with school district prekindergarten programs.

3.4 Early Childhood Mental Health Consultation: Kindergarten – Grade 3

Ten of the 24 grantees (42%) implemented early childhood mental health consultation in elementary schools. One grantee implemented two different approaches with different schools, resulting in a total of 11 programs. Similar to the mental health consultation in early childhood, the mental health consultation in schools involved multiple approaches. The clinicians provided general consultation to teachers and/or counselors on topics such as typical and atypical child development, social-emotional development, or classroom environments that promote the development of social and emotional skills. This type of consultation was not targeted to specific characteristics of the setting but offered mental health-related information that was broadly applicable. In five programs, consultants also provided program-specific consultation on programmatic issues such as programming and curriculum related to positive social and emotional behavior for children. More targeted activities in the schools included child-specific consultation with staff about individual children in the program with behavior concerns. The consultant might observe or assess the child, might meet with the child's parents as well as teachers, suggest strategies or make recommendations to address issues (either specific to a child or to support a positive learning environment that promotes healthy social-emotional development), and could make referrals for additional evaluation or services. Finally, in two programs, consultation activities involved short-term mental health treatment for a child or a parent-child pair. One of the grantees used the Georgetown model for their mental health consultation, but the other grantees developed their own models for delivering consultation in schools.

3.5 Integration of Behavioral Health in Primary Care

Eleven (46%) of the 24 LAUNCH grantees implemented programs to integrate mental/behavioral health into primary care. All of these grantees were in Cohorts 1 or 2. Most of the integration models (75%) involve the physical co-location of the LAUNCH-supported mental health staff in community health care settings. In nine of the programs, LAUNCH mental health staff conducted follow-up assessments of children who were identified as at risk, based on routine screening as part of the visit to the doctor. The mental health staff also met with the medical staff and with the families to discuss the result of assessments during the visit to the pediatrician and to determine an appropriate follow-up plan. All of the grantees provided training to the staff in the health care settings as part of the integration models on topics such as appropriate referrals for children with behavioral/mental health concerns, resources in the community for children with mental/behavioral

health concerns, and strategies for family engagement and working with families to help them understand/support children's healthy development. A small number of grantees provided support for short-term mental health treatment conducted by trained clinicians. These clinicians were available to work with parents and/or children who, through screening and assessment initiated as part of one or more program stategies under LAUNCH, were identified as having mental or behavioral health concerns.

3.6 Systems Change Activities

One of the pillars of the LAUNCH model was a dual focus on systems development as well as service delivery. Among other goals, LAUNCH grantees were expected to help build collaborative relationships among provider organizations across disciplines or systems. To this end, grantees also engaged in activities to enhance the state, tribal, and community early childhood delivery systems and the legislative and organizational policies and practices that influence children's developmental and health outcomes. Grantees implemented six types of systems change activities: partnership development; policy/infrastructure development; data and information systems development; developmental screening/assessment at a population level; workforce development; and public awareness. All grantees were involved in at least one type of systems activity at the state, tribal, and community levels. The majority (83%) of grantees engaged in three or more types of systems activities at the community level. Sixty-one percent (61%) of grantees in Cohorts 1 and 2 implemented three or more types of systems activities at the state level.[13]

One type of partnership development activity was the formation of a Community Young Child Wellness Council at the state, tribal, and community levels. The Community Council was expected to be made up of representatives from multiple agencies and sectors that serve young children and their families—e.g., health (including representatives from the private sector), mental health, child welfare, substance abuse prevention, early childhood education, and local education agencies (Head Start, Early Head Start and Part C), and family representatives (SAMHSA, 2008; 2009; 2010). Grantees in Cohorts 1 and 2 also were expected to form State Young Child Wellness Councils made up of state agencies working on child mental health. The State Councils were to include representatives from health (including representatives from the private sector), mental health, child welfare, Medicaid, substance abuse prevention, early childhood and state education (Early Head Start, Head Start, and Part C), the child care accrediting agency, Title V administering agencies (if applicable), the office of the governor or chief executive of the state/tribe, and families in the target population (SAMHSA, 2008; 2009).

All grantees established Young Child Wellness Councils at the community level. Although each Council established its own priorities for targets of systems change activities, all grantees shared the goal of improving coordination and collaboration among the child and family service providers in the

[13] The Project LAUNCH grants awarded in Cohort 3 were to community-based entities; they were not awarded to the state maternal and child health agency, as in the previous cohorts. Although some Cohort 3 grantees did collaborate with state agencies on early childhood initiatives, this was not as much a focus of their LAUNCH grant as it was in the previous two cohorts.

community. All 18 grantees in Cohorts 1 and 2 established Young Child Wellness Councils at the state level. Although each council established its own priorities for targets of systems change activities, all grantees shared the goal of improving coordination and collaboration among the child and family service providers in the community.

4. FINDINGS ON LAUNCH OUTCOMES: REPRESENTATIVENESS AND STRENGTH OF THE EVIDENCE

As discussed in Chapter 2, the inferences we can draw about LAUNCH from the Cross-Site Evaluation rests in part on the extent to which the findings represent the array of enhancements that LAUNCH grantees have implemented for services and systems and in part on the strength of the evidence from the local evaluations about LAUNCH effects. This chapter summarizes the state of the data on LAUNCH effects, in order to establish the causal inferences we can draw from the existing outcome data about the effectiveness of LAUNCH in achieving its objectives for providers, parents, children, and systems.

4.1 Representativeness of LAUNCH Activities in the Outcome Data

The first step in the meta-analysis was to identify and extract outcome findings from the local grantee evaluations. As discussed earlier, the national evaluation of the effects of Project LAUNCH is intended to report on the effects of the full set of activities that LAUNCH grantees implemented to enhance services in their communities and improve the child services system locally and at the state-level. LAUNCH grantees are expected to work on all of the major components of the local child and family services system (e.g., health care, early childhood education and care, family programs) and on promoting change in policies and practices in the state, tribal, and local service systems. To the extent that the local evaluations reported on outcomes of all of these activities, the conclusions of the meta-analysis represent a more robust estimate of the effectiveness of the LAUNCH approach.

There are at least three important factors that might influence the breadth of the local grantee evaluations. First, in cases where the LAUNCH grantee introduced a new evidence-based program, as opposed to introducing an enhancement like mental health consultation, it would not be expected that the local evaluation would implement an evaluation to collect outcome data on a program that already has been proven to be effective. Second, although the national evaluation would benefit from having data on as many outcomes as possible, the local evaluations had as their primary objective fulfilling the information needs of the grantee. This means that local evaluations might focus on specific new services or enhancements that the grantee wanted to learn about especially, versus covering all of the grantee activities in equal depth. Third, there are pragmatic funding limitations for the local evaluations. Designing and implementing rigorous evaluations of all of a grantee's activities would have been challenging, especially within the budgets typically available to the local evaluators.

The available outcome data from the local evaluations can be compared with the full set of services and systems work being implemented by the LAUNCH grantees. Nearly all grantees implemented some systems change activities. Ten of the grantees (42%) reported on systems outcomes at the local level and two of the eighteen grants with state partners (11%) reported on systems changes at

the state level (Exhibit 6). Most grantees supported service activities in at least three of the four key LAUNCH strategies (home visiting, family strengthening, early childhood mental health consultation, and integration of behavioral health in primary care). All but one grantee reported at at least one outcome finding[14] for one or more LAUNCH-supported services; sixteen of the grantees reported outcome data on two or three key LAUNCH strategies (Exhibit 6). This means that nearly half of the services being supported by LAUNCH grantees are represented in the outcome analysis. However, this count includes provider, parent, and child outcomes. For any single one of these respondent groups, the number of services for which outcomes reported is much smaller. As a result, the meta-analysis of effects are based on a non-random subset of the LAUNCH services, which urges caution in generalizing the findings to LAUNCH as a whole.

Exhibit 6. Representation of Grantees and Programs in the Outcome Analysis

Cohort	Systems Outcomes	Services Outcomes
Cohort 1 (n = 6 grantees)	None: 5 grantees Local: 1 grantee State: 0 grantees	None: 0 grantees 1 strategy: 1 grantee 2 strategy: 4 grantees 3+ strategy: 1 grantees
Cohort 2 (n = 12 grantees)	None: 6 grantees Local systems: 4 grantees State & local systems: 2 grantees	None: 1 grantee 1 strategy: 5 grantees 2 strategy: 6 grantees 3+ strategy: 0 grantees
Cohort 3 (n = 6 grantees)	None: 1 grantee Local: 5 grantees State: NA	None: 0 grantees 1 strategy: 1 grantee 2 strategy: 5 grantees 3+ strategy: 0 grantees

Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

When the outcome data are organized by the target sample (providers, parents, or children) and by type of program or program strategy (home visiting, family strengthening, early childhood mental health consultation, or integration of behavioral health in primary care), the small number of outcomes is evident (Exhibit 7). For example, child outcomes are reported for between 5 percent and 45 percent of the programs in each strategy. For parent outcomes, the range is 0 percent to 31 percent. The uneven distribution of the outcome data and the limited representation of the full set of LAUNCH programs is a reason for exercising caution in interpreting the results of the analyses as providing a valid summary of the effectiveness of Project LAUNCH as a whole.

4.2 Strength of the Evidence on LAUNCH Outcomes

As described earlier, the R-SEED framework was used to assess the strength of the evidence on LAUNCH outcomes. The strength of evidence establishes the confidence with which we attribute a causal link between the LAUNCH activities and gains shown by providers, parents, or children.

[14] This includes "codable" outcomes, where "codable" means data for which we can create a standardized effect size for the meta-analysis.

That is, for effects that have been generated by studies with less rigor, we have less confidence that LAUNCH is the causal agent responsible for those effects. Lower rigor studies do not allow us to eliminate other possible causes for outcomes, such as normal development (for child outcomes), time, or history (e.g., other events that could affect outcomes for members of a community).

4.2.1 Evidence on Provider Outcomes

A small number of provider effects received R-SEED ratings of Emerging or Strong evidence (Exhibit 8). This stronger evidence was reported for the mental health consultation programs in preschool care settings, where the provider outcomes were measured by standardized classroom environment measures. In the other strategies, all of the contrasts were assigned R-SEED evidence ratings of Limited with Reservations or Weak, primarily because of the lack of standardized measures of relevant provider outcomes.

Exhibit 7. Proportion of LAUNCH-Supported Programs Contributing to the Outcome Analysis

Strategy	# of grantees implementing strategy (out of 24)	# of LAUNCH-supported programs across 24 grantees	# Grantees Reporting Outcomes/% of Grantees Implementing Strategy			# Programs Reporting Outcomes/ % of Programs in Strategy		
			Provider Outcomes	Parent Out-comes	Child Out-comes	Provider Out-comes	Parent Out-comes	Child Out-comes
Home visiting	21	26	1 (5%)	3 (14%)	4 (19%)	1 (4%)	4 (15%)	4 (19%)
Family strengthening	22	52	2 (9%)	15 (68%)	3 (14%)	2 (4%)	16 (31%)	3 (5%)
Early childhood mental health consultation: Preschool	18	20	7 (39%)	0	6 (33%)	10 (50%)	0	8 (45%)
Early childhood mental health consultation: K–Grade 3	10	11	0	0	4 (40%)	0	0	4 (36%)
Integration of behavioral health in primary care	11	11	2 (18%)	2 (18%)	2 (18%)	2 (18%)	2 (18%)	2 (18%)

Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

4.2.2 Evidence on Parent Outcomes

Nearly all of the effects on parents were generated by pre-post studies using non-standardized measures, which is reflected in their preponderance of contrasts receiving R-SEED ratings of Weak

evidence or evidence that is Limited with Reservations (Exhibit 9). The strategy with stronger evidence on parent outcomes is Family Strengthening. These higher evidence ratings are related primarily to the fact that more of the parent outcome measures were normed, standardized measures. One reason is that the multiple Incredible Years programs in the sample all used the same two standardized parent outcome measures.

Exhibit 8. Number of Provider Effects by Strength of Evidence for LAUNCH-Supported Services by Strategy

	Strength of Evidence (R-SEED Rating)					
	Weak	Limited with Reservations	Limited	Emerging	Strong	All Evidence (# grantees/ # programs)
Home Visiting						
# contrasts	0	7	0	0	0	7 (1/1)
Family Strengthening						
# contrasts	16	0	0	0	0	16 (2/2)
Early Childhood Mental Health Consultation: Preschool						
# contrasts	15	113	11	25	5	169 (7/10)
Early Childhood Mental Health Consultation: Kindergarten: Grade 3						
# contrasts	0	0	0	0	0	0
Integration of Behavioral Health in Primary Care						
# contrasts	4	10	0	0	0	14 (2/2)
All Strategies						
# contrasts (% total contrasts)	35 (14%)	130 (69%)	11 (4%)	25 (10%)	5 (2%)	206 (11/14)

Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

Exhibit 9. Number of Parent Effects by Strength of Evidence for LAUNCH-Supported Services by Strategy

	Strength of Evidence (R-SEED Rating)					
	Weak	Limited with Reservations	Limited	Emerging	Strong	All Evidence (# grantees/ # programs)
Home Visiting						
# contrasts	26	97	0	0	0	123 (3/4)
Family Strengthening						
# contrasts	45	45	59	3	1	153 (15/16)
Early Childhood Mental Health Consultation: Preschool						
# contrasts	0	0	0	0	0	0
Early Childhood Mental Health Consultation: Kindergarten: Grade 3						
# contrasts	0	0	0	0	0	0
Integration of Behavioral Health in Primary Care						
# contrasts	0	10	12	0	0	22 (2/2)
ALL STRATEGIES						
# contrasts (% total contrasts)	71 (25%)	152 (49%)	71 (25%)	3 (1%)	1 (< 1%)	298 (18/22)

Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

4.2.3 Evidence on Child Outcomes

The child outcomes had higher strength of evidence ratings, compared to the provider and parent outcomes. Whereas the majority of parent and provider outcomes were assigned the lowest two R-SEED ratings, the comparable percentage for child outcomes was 20 percent (Exhibit 10). This is primarily because of the availability of standardized child measures in most developmental domains; 80 percent of the contrasts for child outcomes had an R-SEED rating of Limited, Emerging, or Strong, although most fell in the Limited category.

Exhibit 10. Number of Child Effects by Strength of Evidence for LAUNCH-Supported Services by Strategy

	Strength of Evidence (R-SEED Rating)					
	Weak	Limited with Reservations	Limited	Emerging	Strong	All Evidence (# grantees/ # programs)
Home Visiting						
# contrasts	1	31	16	0	0	48 (4/5)
Family Strengthening						
# contrasts	1	3	9	0	0	13 (3/3)
Early childhood Mental Health Consultation: Preschool						
# contrasts	1	0	54	0	0	55 (9/6)
Early Childhood Mental Health Consultation: Kindergarten: Grade 3						
# contrasts	0	0	59	2	3	64 (4/4)
Integration of Behavioral Health in Primary Care						
# contrasts	0	8	30	0	0	38 (2/2)
All Strategies						
# contrasts (% total contrasts)	3 (1%)	42 (19%)	168 (78%)	2 (1%)	3 (1%)	218 (13/14)

Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

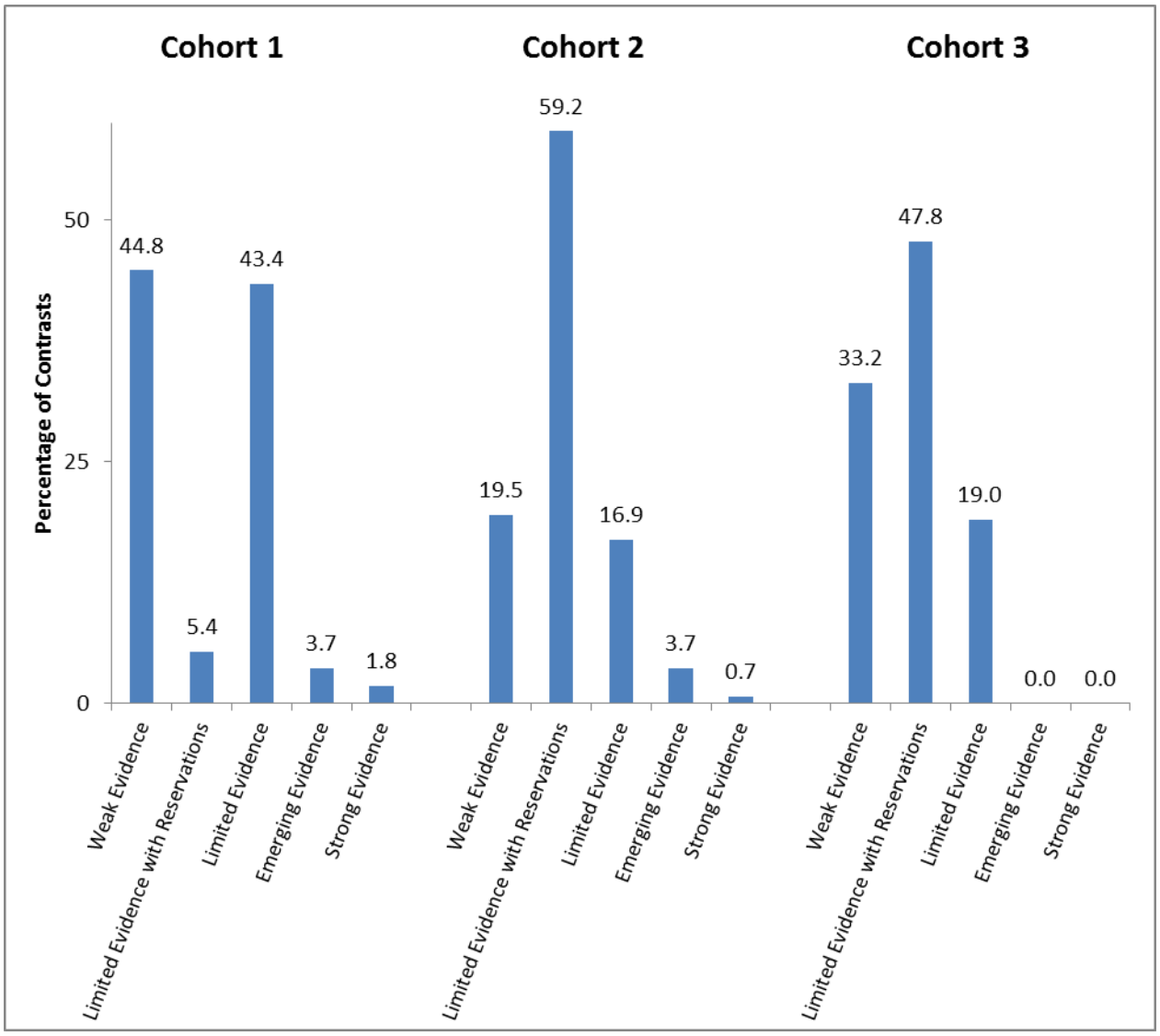
4.2.4 Strength of Evidence by Cohort

The Cross-Site Evaluation examined variation in the strength of evidence of the outcomes findings across the three cohorts. One possible hypothesis is that strength of evidence would be highest for Cohort 1, since these grants had been operating longer and the evaluators had had more time with the programs. There was evidence to support this hypothesis (Exhibit 11). On average, more of the contrasts from the Cohort 1 evaluations had higher ratings for strength of evidence, compared with either the Cohort 2 or the Cohort 3 findings.

4.2.5 Evidence on Systems Changes

All of the evidence on systems changes was rated as either Weak or Limited with Reservations (Exhibit 12). These ratings reflect the fact that most of the local evaluations designed pre-post studies of changes in collaboration among agencies and organizations in the child service system using the Wilder Collaboration Factors Inventory or a similar type of survey. None of the studies included more than two baseline measures of collaboration and none used a normed measure.

Exhibit 11. Strength of Evidence of Effects by Cohort



Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

Although systems initiatives also included policy development, public awareness, workforce development, developmental screening, and data systems enhancements, grantees did not conduct rigorous evaluations of these other activities. Designing evaluations of systems change is challenging for a number of reasons. First, there may be different perspectives on what constitutes "success" or even an "improvement" over a baseline condition. Second, it is often difficult to find an appropriate counterfactual against which to measure the impact of systems change (e.g., a new policy). For example, an appropriate measure of the impact of a new policy or policy change would be where the state or community is now compared to where it would be if it had continued without the policy reform. There can be disagreement over what would have happened if a systems change

had not occurred. Third, it can be hard to link causation to a particular initiative. When the external environment is dynamic, systems change can be influenced by a variety of factors or initiatives other than the program (e.g., Project LAUNCH) being studied.

Exhibit 12. Number of Systems Change Effects by Strength of Evidence

	Strength of Evidence (R-SEED Rating)					
	Weak	Limited with Reservations	Limited	Emerging	Strong	All
Local						
# contrasts (# grantees)	134	131	--	--	--	265 (9)
State						
# contrasts (# grantees)	58	60	--	--	--	118 (2)
All						
# contrasts (% total contrasts)	192 (50%)	191 (50%)	--	--	--	383

Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013, and (ii) Annual program reports 2009 – 2013.

5. EFFECTS OF LAUNCH PREVENTION AND PROMOTION STRATEGIES

This section presents the results of the analyses of the Project LAUNCH outcomes of providers, parents and children resulting from five LAUNCH prevention and promotion strategies. Separately, for each group of individuals, these analyses look at relevant outcomes for:

- LAUNCH-enhanced home visiting programs;
- LAUNCH-enhanced family strengthening programs;
- LAUNCH early childhood mental health consultation programs in preschool settings (primarily provider and child outcomes);
- LAUNCH early childhood mental health consultation programs in kindergarten through grade 3 (primarily provider and child outcomes); and
- LAUNCH programs to integrate behavioral health in primary care (primarily provider outcomes).

The section also reports on systems outcomes of LAUNCH activities to improve the child and family services system, at both the local, tribal, and state levels.

5.1 Overall Effects

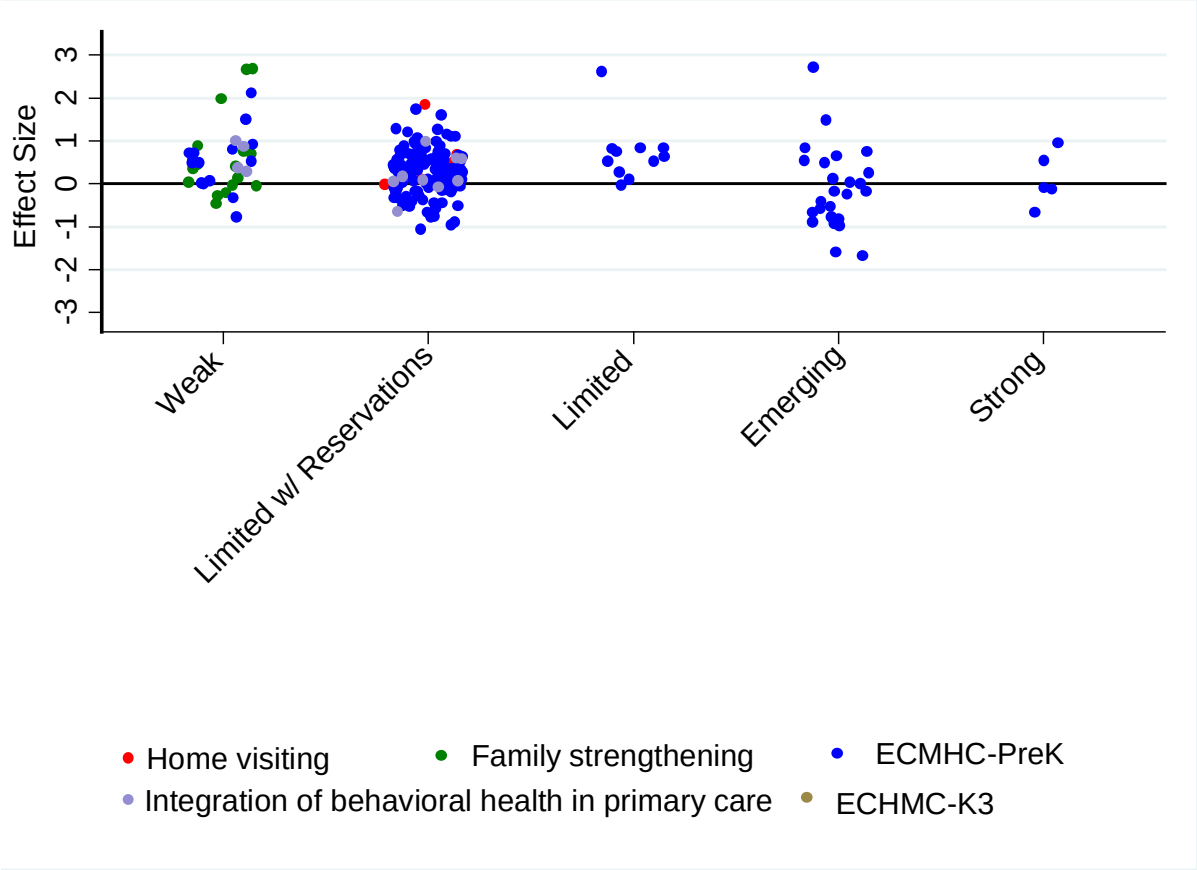
The effects described in this section are organized by respondent—provider, parent, or child—within strategy. For each respondent type, one set of analyses examines the average effect sizes weighted by precision. A second set of analyses adds weights for the strength of evidence.

5.1.1 Effects on Providers

Across all program strategies and all strengths of evidence, the overall average effect on providers is small-to-moderate (ES = 0.38) (Exhibit 13). The average effect ranges from a high of 0.55 for the effects assessed as providing Weak evidence to a low of 0.15 for the small number of effects rated as providing Strong evidence (Figure 2). This trend is consistent with other meta-analyses that report larger average effects for weaker studies. When the overall average is adjusted for strength of evidence, the average is decreased because of the "down-weighting" of the effects at the lowest evidence ratings. The overall average adjusted for strength of evidence is ES = 0.32. This finding suggests that if the meta-analysis only included studies with more rigorous designs, the effects would be smaller.

The same pattern is found within program. The type of program with the largest number of provider effects was mental health consultation in early care and education settings. For these programs, the overall effect on providers prior to adjusting for strength of evidence was ES = 0.34. When the effects are weighted by R-SEED rating, the average effect on providers drops (ES = 0.28).

Figure 2: Estimated Effects of LAUNCH on Providers by Strength of Evidence Rating and Strategy



Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

Exhibit 13. Average Effect Sizes[a] for Provider Outcomes of LAUNCH-Supported Services by Strength of Evidence

	Strength of Evidence (R-SEED Rating)						
	Weak	Limited with Reservations	Limited	Emerging	Strong	All Evidence	All Evidence Adjusted for Strength of Evidence[b]
Home Visiting							
Average effect size (standard error)	--	0.54	--	--	--	0.54 (0.14)	0.54 (0.14)
Family Strengthening							
Average effect size (standard error)	0.85	--	--	--	--	0.85 (0.22)	0.85 (0.22)
Early Childhood Mental Health Consultation: Preschool							
Average effect size (standard error)	0.47	0.28	0.47	0.03	0.15	0.34 (0.03)	0.28 (0.04)
Early Childhood Mental Health Consultation: Kindergarten – Grade 3							
Average effect size (standard error)	--	--	--	--	--	--	--
Integration of Behavioral Health in Primary Care							
Average effect size (standard error)	0.72	0.22	--	--	--	0.28 (0.12)	0.25 (0.11)
ALL PROGRAM STRATEGIES							
Average effect size (standard error)	0.55	0.35	0.47	0.03	0.15	0.38 (0.03)	0.32 (0.04)

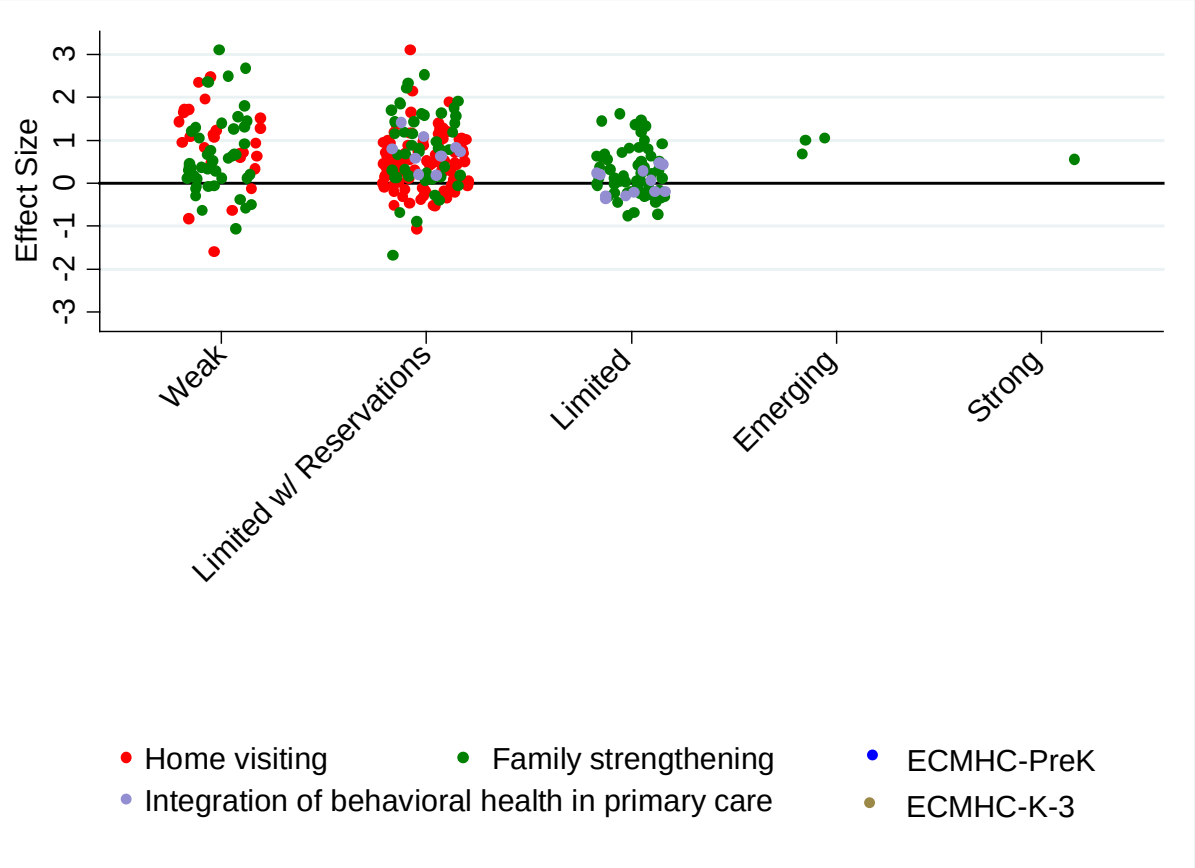
[a] Effect sizes weighted for precision of estimate.
[b] Effect sizes weighted by R-SEED rating

Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

5.1.2 *Effects on Parents*

Across all program strategies, the overall average effect on parent outcomes was small-to-moderate (ES = 0.48) when the effects were not weighted by strength of evidence (Exhibit 14). Among these programs, the average effect was highest for the effects assessed as providing Strong evidence (ES = 0.71) (Figure 3). However, when the overall average is adjusted for strength of evidence, it is lower (ES = 0.44); the number of contrasts rated as Strong evidence was so small that the average was still reduced because of the lower ratings for the more predominant weaker evidence.

Figure 3: Estimated Effects of LAUNCH on Parents by Strength of Evidence Rating and Strategy



Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

The average effects on parents vary by strategy (Exhibit 14). After adjusting for strength of evidence, the average effect on parents is small for two strategies--family strengthening and integration of behavioral health in primary care (ES = 0.31 and 0.20, respectively). In home visiting, there is a large average effect (ES = 0.71), but this is based on a small number of contrasts rated as evidence that is Weak and Limited with Reservations, so the average has to be viewed very cautiously.

Exhibit 14. Average Effect Sizes[a] for Parent Outcomes of LAUNCH-Supported Services by Strength of Evidence

Contrast-level effect size	Strength of Evidence (R-SEED rating)						
	Weak	Limited with Reservations	Limited	Emerging	Strong	All Evidence	All Evidence Adjusted for Strength of Evidence[b]
Home Visiting							
Average effect size (standard error)	0.71	0.71	--	--	--	0.71 (0.05)	0.71 (0.05)
Family Strengthening							
Average effect size (standard error)	0.39	0.28	0.34	0.73	0.71	0.31 (0.03)	0.31 (0.04)
Early Childhood Mental Health Consultation: Preschool							
Average effect size (standard error)	--	--	--	--	--	--	--
Early Childhood Mental Health Consultation: Kindergarten - Grade 3							
Average effect size (standard error)	--	--	--	--	--	--	--
Integration of Behavioral Health in Primary Care							
Average effect size (standard error)	--	0.65	0.16	--	--	023 (0.06)	0.20 (0.05)
ALL PROGRAM STRATEGIES							
Average effect size (standard error)	0.51	0.54	0.29	0.73	0.71	0.48 (0.03)	0.44 (0.03)

[a] Effect sizes weighted for precision of estimate.

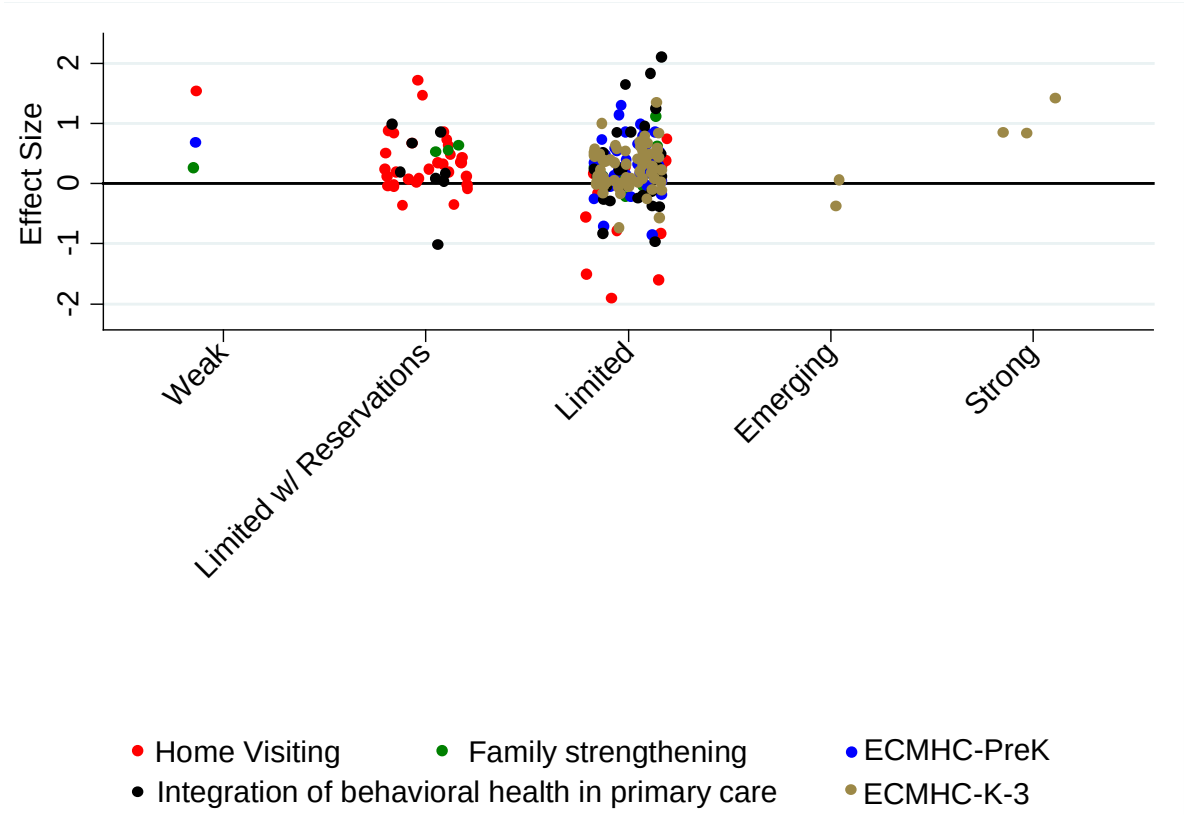
[b] Effect sizes weighted by R-SEED rating

Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

5.1.3 *Effects on Children*

Across all program strategies, the average effect on children was small (ES = 0.28) (Exhibit 15). The size of the overall effects range from a high of 0.81 for the effects assessed as providing Weak evidence to a low of 0.25 for the effects rated as providing Limited evidence (Figure 4). When the overall average is adjusted for strength of evidence, it is virtually identical (ES = 0.29).

Figure 4: Estimated Effects of LAUNCH on Children by Strength of Evidence Rating and Strategy



Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

The average effects on children, after adjusting for strength of evidence, are similar across LAUNCH strategies and small (Exhibit 15). Across the strategies, the average effect on children ranges from 0.24 to 0.31. None of the strategies had a large number of child effects reported.

Exhibit 15. Average Effect Sizes[a] for Child Outcomes of LAUNCH-Supported Services by Strength of Evidence

Contrast-level effect size	Strength of Evidence (R-SEED rating)						
	Weak	Limited with Reservations	Limited	Emerging	Strong	All Evidence	All Evidence Adjusted for Strength of Evidence[b]
Home Visiting							
Average effect size (standard error)	1.61	0.23	-0.17	--	--	0.26 (0.07)	0.20 (0.06)
Family Strengthening							
Average effect size (standard error)	0.30	0.68	0.21	--	--	0.38 (0.13)	0.32 (0.15)
Early Childhood Mental Health Consultation: Preschool							
Average effect size (standard error)	0.40	--	0.27	--	--	0.27 (0.03)	0.27 (0.03)
Early Childhood Mental Health Consultation: Kindergarten - Grade 3							
Average effect size (standard error)	--	--	0.24	0.07	0.61	0.26 (0.04)	0.31 (0.04)
Integration of Behavioral Health in Primary Care							
Average effect size (standard error)	--	0.38	0.22	--	--	0.24 (0.04)	0.23 (0.04)
ALL PROGRAM STRATEGIES							
Average effect size (standard error)	0.81	0.29	0.25	0.36	0.61	0.28 (0.02)	0.29 (0.02)

[a] Effect sizes weighted for precision of estimate.
[b] Effect sizes weighted by R-SEED rating.

Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

5.1.4 Average Effects on Systems

The average effects on local systems change was small (ES = 0.32) (Exhibit 16). When the effects were adjusted for their strength of evidence, the average drops (ES = 0.27). The average effect on

state systems change was moderate in size (ES = 0.50). This average also was smaller when effects were weighted by strength of evidence (ES = 0.41).

All of the evidence on systems change was rated as providing Weak evidence or evidence that is Limited with Reservations. The larger effect sizes for systems change is very likely related to the fact that the studies contributing evidence were of low rigor and therefore more likely to have bigger effect sizes.

Exhibit 16. Average Effect Sizes [a] for Systems Change Outcomes for LAUNCH Grantees by Strength of Evidence

Contrast-level effect size	Strength of Evidence (R-SEED rating)						
	Weak	Limited with Reservations	Limited	Emerging	Strong	All Evidence	All Evidence Adjusted for Strength of Evidence b
Local							
Average effect size (standard error)	0.49	0.19	--	--	--	0.32 (0.19)	0.27 (0.02)
State							
Average effect size (standard error)	0.80	0.25	--	--	--	0.50 (0.41)	0.41 (0.04)

[a] Effect sizes weighted for precision of estimate.
[b] Effect sizes weighted by R-SEED rating.

Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013, and (ii) Annual program reports 2009 – 2013.

5.1.5 Interpreting Effects Obtained for LAUNCH

The meta-analysis indicated that LAUNCH had overall positive effects. The framework underlying LAUNCH hypothesized that there would be positive effects on children to the extent that the LAUNCH-supported changes to services and systems were associated with changes in the providers of services for parents and children and changes in parents themselves. The overall effects for providers and parents were larger, on average, than the effects for children. The fact that effects on all three groups remain after being adjusted for strength of evidence allows us to conclude that they are experiencing positive growth during the time LAUNCH is supporting programs and systems changes. The adjustments for strength of evidence do not, however, address the fact that in the absence of more rigorous studies that employ appropriate comparison groups, we cannot say that this growth is solely the result of LAUNCH (i.e., would not have happened in the absence of LAUNCH).

How can we interpret the size of the effects obtained in these analyses? One of the metrics to help us understand the LAUNCH effects is other meta-analytic work looking at the effects of relevant interventions. Relevant meta-analyses have been conducted for all of the LAUNCH strategies, but not always for all of the LAUNCH respondents—providers, parents, and children. It is important to

underscore that published meta-analyses limit their database to studies using either randomized or quasi-experimental designs. In the LAUNCH data, these designs would correspond to Strong evidence, and it has already been noted that there are just a few LAUNCH effects based on these designs. It is known that effects generated by less rigorous designs tend to be larger, so comparisons of average effects using the full set of LAUNCH data can be expected to show LAUNCH effects as being larger.

In general, the LAUNCH effects are in the same range as many of the effects reported in other research (Exhibit 17) (Sweet and Appelbaum, 2004; Nievar, 2009; Layzer et al., 2001; Raver et al., 2008; ED, 2011).

- For the child outcomes, the magnitude of the effects obtained in the LAUNCH analyses was very similar to the size of the effects reported in other meta-analyses.
- For parent outcomes, the research is more mixed in the overall effects reported, so the comparison with LAUNCH is less clear. The LAUNCH effects for home visiting programs are larger than have been reported in some other research but quite similar to the effects in other studies.

Exhibit 17. Comparison of LAUNCH Effects and Effects in Related Meta-Analyses

	Providers		Parents		Children	
	LAUNCH (All levels of evidence)	Other Reviews (Strong evidence)	LAUNCH (All levels of evidence)	Other Reviews (Strong evidence)	LAUNCH (All levels of evidence)	Other Reviews (Strong evidence)
Home visiting	0.54 (0.14)		0.71 (0.05)	.14 (.04)[15]	0.20 (0.06)	0.37 (N/A)[16]
Family Strengthening	0.85 (0.22)		0.32 (0.04)	.26 (.03)[17]	0.32 (0.15)	0.22 (.03)[18]
Early Childhood Mental Health Consultation: Prek-socio-emotional curricula	0.28 (0.04)	0.30 - 0.53 for different curricula (N/A)[19]			0.27 (0.03)	0.46 (N/A)[20]
Early Childhood Mental Health Consultation: K to Grade 3--socio-emotional curricula					0.31 (0.04)	0.30 (N/A)[21]

15 Sweet, M.A. and Appelbaum, M.I. (2004)
16 Nievar (2009)
17 Layzer, J. I., Goodson, B. D., Bernstein, L., & Price, C. (2001)
18 Layzer, J. I., Goodson, B. D., Bernstein, L., & Price, C. (2001)
19 Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K. M., & Sardin, L. (2008)
20 U.S. Department of Education, Institute of Education Sciences What Works Clearinghouse (2011)
21 Durlak(2011).

Only one prior study examined provider outcomes, in the context of early childhood curricula focused on socio-emotional development. The effect sizes for changes in provider practice reported in prior research ranged from ES = 0.30 to ES = 0.50, and the parallel LAUNCH effect for providers was 0.28 to 0.54. The fact that the average LAUNCH effects, adjusted for strength of evidence, are comparable to many other meta-analytic findings, increases our confidence that the LAUNCH results can be seen as reasonable, i.e., not badly skewed by the weaker evidence. At the same time, it is critical to remember that because the research in the outside analyses are based on strong (i.e., internally valid) designs, the effects reported can be attributed to the interventions being studied; for LAUNCH, this attribution cannot be made with confidence.

5.1.6 Variation in LAUNCH Outcomes

The previous section presented the overall effects on providers, parents, and children for key LAUNCH promotion and prevention strategies and for local and state systems changes. Additional analyses explored relationships between effects and a small set of implementation factors. The implementation factors are listed below and defined in more detail in the subsequent discussion:

- Program-level breadth of LAUNCH activities to integrate mental and behavioral health into services;
- Program-level mental health consultation on classroom environment in early childhood and care settings;
- Grantee-level breadth of LAUNCH activities to improve state and local child service systems; and
- Cohort in which grantee was funded.

The relational analyses examine differences in the average outcomes for providers, parents, and children, as a function of these implementation factors, across and within program strategies.

5.1.7 Implementation Factors in the Relational Analyses

5.1.7.1 LAUNCH approaches to integration of mental/behavioral health into services

As described earlier, Project LAUNCH grantees were expected to enhance the child services system by filling service gaps, implementing workforce development around child mental health and assessment, and integrating awareness of maternal and child mental and behavioral health into services. These goals were common across the four prevention and promotion strategies, although the ways in which mental health integration was implemented varied depending on the type of program involved. Two measures of implementation of mental health consultation were created. These scores were entered into analyses of the variation in average effects on providers, parents and children.

One measure described, for each LAUNCH-supported program, the breadth of the LAUNCH-supported activites to integrate child mental health into services. Each program could receive a point (Yes/No) for each of nine integration activities (see Box C). The total score for Breadth of Strategies for Mental Health Consultation was entered in analyses of the variation in average effects on providers, parents and children. Across all program strategies, the average score was 3.5 out of 9 (Exhibit 18). The average score varied by strategy: The lowest average score on breadth of activities

was for family strengthening programs and the highest scores were for early childhood mental health consultation and integration of behavioral health in primary care. The average score on classroom consultation was 1.4 out of 3 for preschool settings and 1.5 out of 3 for elementary school settings.

Box C
Implementation Factors in Relational Analyses of LAUNCH Outcomes

- Breadth of LAUNCH activities for integration of mental and behavioral health into services (0-9)
  - Trained mental health consultant funded to work with program
  - Mental health consultant provides staff training on child/parent mental/behavioral health issues
  - Mental health consultant provides staff training on assessment of all domains of child development
  - Grantee funds staff assessment of all domains of child development
  - Mental health consultation introduces new programming/curriculum related to child mental/behavioral health
  - Mental health consultant discusses individual parents/children with mental/behavioral health concerns
  - Mental health consultation Is guided by/based on theoretical framework (e.g., CSEFEL, PBIS)
  - Mental health consultant provides clinical supervision of staff
  - Mental health consultant provides brief individual therapy to parents/children with mental/behavioral health concerns
- Mental health consultation on classroom environment in early childhood settings (0-3)
  - Mental health consultant provides consultation on classroom environment
  - Mental health consultant meets with staff on a regular, scheduled basis (vs as-needed)
  - Mental health consultant is embedded in program on full-time basis
- Breadth of LAUNCH activities to improve state and local child service systems: # types of local/state systems initiatives (0 – 6 for local, 0-6 for state, 0-12 overall)
  - Development of coordination of agencies, departments involved in child and family services
  - Initiatives related to policy or infrastructure development around child mental health
  - Initiatives around integrated data (across providers, agencies) on child and family mental/behavioral health Data
  - Policies to promote screening/assessment of child and family mental/behavioral health
  - Enhancement of workforce knowledge/practices related to child and family mental/behavioral health
  - Public awareness of child and family mental/behavioral health
- Cohort (1-3)
  - Funding cohort 1: 2008-09
  - Funding cohort 2: 2009-10
  - Funding cohort 3: 2010-11

Exhibit 18. Summary Statistics for Implementation Factor Scores for Breadth of Activities to Integrate Mental and Behavioral Health into Services and Consultation on Classroom Environment in Early Childhood Education and Care Settings

Promotion and Prevention Strategy	Breadth of Strategies in Mental Health Consultation		Mental Health Consultation on Classroom Environment	
	Mean (s.d.)	Range (out of 9)	Mean (s.d.)	Range (out of 3)
Home visiting	3.0 (2.4)	(1-7)		
Family strengthening	2.5 (1.4)	(1-5)		
Early childhood mental health consultation: PreK	4.8 (1.8)	(2-7)	1.4 (0.7)	(0-3)
Early child mental health consultation: K – grade 3	3.7 (1.5)	(2-5)	1.5 (0.5)	(1-2)
Integration of behavioral health in primary care	4.7 (0.5)	(4-5)		
All strategies	3.5 (1.8)	1-7		

Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013.

5.1.7.2 Breadth of LAUNCH activities to improve state and local child service systems

This implementation variable is defined as the number of categories of systems change in which grantees undertook initiatives in the LAUNCH community (local) and at the state level. (Although the Cohort 3 grantees did not have a state partner, some of the grantees implemented initiatives at the state level.) Grantees were scored (0/1) for each of six categories of systems-change initiatives (see Box C). Grantees were given a score if they undertook one or more initiatives in a domain, i.e., the indicator is about number of domains in which grantees worked, not on number of different initiatives undertaken. An overall score was calculated by summing the scores at the local and state levels; the total score ranged from 0 to 6 for local and state initiatives respectively, and 0 to 12 for the overall score.

At the local level, the average number of domains in which grantees implemented systems change initiatives was 4.1 (out of 6), with the scores ranging from 1 to 5 (Exhibit 19). At the state level, the average number of domains of state-level initiatives was smaller—2.8 (out of 6). The combined total across local and state domains was an average of 6.9. Across the 23 grantees in the analysis sample[22], 43 percent had a total score of 8 – 10 domains; 47 percent had between 5 and 7 domains; and the remaining 10 percent had fewer than 5 domains.

[22] One grantee did not report any outcome data and was dropped from all relational analyses. The analysis sample therefore included 23 grantees.

Exhibit 19. Summary Statistics for Implementation Factor Score for *Breadth of Systems Change Initiatives at the Local and State Levels*

	Systems Change Score	
	Mean (s.d.)	Range (out of 6)
Local systems change initiatives	4.1 (1.3)	(1-5)
State systems change initiatives	2.8 (1.3)	(1-5)
Total systems change initiatives	6.9 (1.9)	(1-12)

Source: Analysis of data from annual end-of-year local evaluation reports 2009 – 2013 and annual program reports 2009 – 2013.

5.1.8 Relationship of Implementation Factors to LAUNCH Outcomes

5.1.8.1 LAUNCH approaches to integration of mental/behavioral health into services

Breadth of LAUNCH activities. It was hypothesized that, within each program strategy, higher scores on breadth of activities to integrate mental/behavioral health into services would be positively related to size of effects. Across the fourteen tests that could be conducted (i.e., where there was sufficient variation in the outcomes to support the analysis), there were four positive and statistically significant relationships between intensity of mental health consultation and program effects: for parent outcomes in home visiting and integration of behavioral health in primary care, for provider outcomes in family strengthening, and for child outcomes in mental health consultation programs in schools (Exhibit 20). Three other tests resulted in findings that were in a positive direction but did not reach statistical significance. Because the number of significant relationships is greater than chance, the analyses suggest that variation in program effects may, in fact, be related to variation in the kinds of supports that LAUNCH grantees have implemented to bring mental and behavioral health into the child and family services systems. The results for the variable measuring classroom consultation in early childhood settings followed the same pattern as the overall findings, with one significant relationship for child outcomes in programs in school settings.

Exhibit 20. Results of Analyses Relating Breadth of Mental Health Integration Activities to Average Effect Size by Strategy and Sample

Strategy and sample	# Programs	# Contrasts	Overall Intensity of Mental Health Integration Size and Direction of Relationship
Home Visiting			
Provider	1	4	NA[b]
Parent	4	123	Significant and positive [p < .001]
Child	5	48	NS
Family Strengthening			
Provider	2	16	Significant and positive [p < .01]
Parent	18	153	NS
Child	3	13	NS
Early Childhood Mental Health Consultation: Preschool			
Provider	12	159	NS
Parent	0	0	NA[b]
Child	10	55	Positive [p < .10]
Early Childhood Mental Health Consultation: Kindergarten – Grade 3			
Provider	0	0	NA[b]
Parent	0	0	NA[b]
Child	6	64	Significant and positive [p < .01]
Integration of Behavioral Health in Primary Care			
Provider	2	14	NS
Parent	2	22	Significant and positive [p < .001]
Child	3	38	Positive [p < .10]
All LAUNCH Service Strategies			
Provider	17	196	NS
Parent	24	298	NS
Child	27	218	Positive [p < .10]

[a] Not tested in this strategy.

[b] Insufficient variation to support analysis model

NS = not-statistically significant at p < .05.

Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

5.1.8.2 *Breadth of LAUNCH activities to improve state, tribal, and local child service systems*

It was hypothesized that the breadth of systems change activities—i.e., the number of categories of systems change in which grantees undertook initiatives in the LAUNCH community and at the state level (see Box C, above)—to improve state, tribal, and local child service systems would have a positive relationship to outcomes. In other words, the more comprehensive grantees were in their systems change efforts, the more likely it would be that they have positive parent, child, and provider outcomes. The analyses confirm this hypothesis: The breadth of systems change activities was positively related to the overall effects on parents and overall effects on the extent of collaboration reported at the state level (Exhibit 21).

Exhibit 21. Results of Analyses Relating Breadth of Systems Change Activities to Average Effect Size by Strategy and Respondent

Strategy and Sample	# Programs	# Contrasts	Significance and Direction of Relationship
Provider			
All outcomes, all strategies	10	209	NS
Parent			
All outcomes, all strategies	15	296	Significant and positive [p < .001]
Child			
All outcomes, all strategies	14	223	NS
Local Systems Changes			
Collaboration	9	265	NS
State Systems Changes			
Collaboration	2	118	Significant and positive [p < .001]

Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014); (iv) Annual program reports 2009 – 2013.

5.1.8.3 *Cohort*

There were six possible tests of the relationship between cohort and average effect size. There was a significant relationship with cohort for three of the six tests (Exhibit 22). In all of these tests, the average effects for Cohort 1 are higher, on average, then outcomes for the other two cohorts. Earlier (see page 33), we identified reasons why Cohort 1 programs would have better outcomes than were found in Cohorts 2 and 3. The primary reason is that the Cohort 1 programs have been in place longer and have had more time for program enhancements and/or programs to mature and become more effective and for parents to have received longer doses of a program's services. If these are reasons for the better performance of Cohort 1 programs, then over time, the effects in the later cohorts should "catch up" when their programs have been in place for more years. The examination of variation in effects focused on a small set of implementation factors—one related to

Exhibit 22. Results of Analyses Relating Cohort to Average Effect Size by Program Strategy and Respondent

Program Strategy and Sample	# Grantees	# Contrasts	Size and Direction of Relationship			Finding
			Cohort 1	Cohort 2	Cohort 3	
Home Visiting						
Provider	1	4	Insufficient variation for model			--
Parent	3	123	NS	Negative [p < .10]	Positive [p < .001]	Cohort 3 > Cohorts1, 2 Cohorts 1, 3 > Cohort 2
Child	4	48	positive [p < .001]	NS	NS	Cohort 1 > Cohorts 2,3
Family Strengthening						
Provider	2	16	Insufficient variation for model			--
Parent	15	153	NS	NS	NS	No significant difference
Child	3	13	Insufficient variation for model			--
Early Childhood Mental Health Consultation: Preschool						
Provider	7	169	NS	Insufficient variation for model		No significant difference
Parent	0	0	NA			--
Child	8	70	Positive [p < .10]	NS	Significant and negative [p < .001]	Cohorts1, 2 > Cohort 3
Early Childhood Mental Health Consultation: Kindergarten - Grade 3						
Provider	0	0	NA			--
Parent	0	0	NA			--
Child	3	64	Insufficient variation for model			--
Integration of Behavioral Health in Primary Care						
Provider	2	14	Insufficient variation for model			--
Parent	1	22	Insufficient variation for model			--
Child	3	38	Insufficient variation for model			--
Local Systems Changes						
Collaboration	9	265	NS	NS	NS	No significant difference
State Systems Changes						
Collaboration	2	118	Insufficient variation for model			--

Source: Analysis of data from (i) Annual end-of-year local evaluation reports 2009 – 2013; (ii) Maine Special Studies Final Report Studies One and Two and Final Report for Study Three (School of Community and Population Health, University of New England, 2014); (iii) interim findings from Red Cliff Special Study (Brazelton Touchpoints Center, 2014).

the approaches to integrating mental and behavioral health in individual LAUNCH programs, another related to the breadth of initiatives to improve the child service system at the local and state levels, and a third related to length of time that grantees had been implementing LAUNCH. There were a

few positive relationships that suggested that the implementation of the LAUNCH model was related to the size of the observed effects. However, the few findings, laid against the total number of relationships tested, and laid against the concerns with the quality of the evidence, means that the observed results have to be seen as suggestive and potentially worth testing in the future as opposed to confirming true relationships.

6. FINDINGS FROM THE SAMHSA PROVIDER AND PARENT SURVEYS

This section of the report presents findings on provider outcomes that were generated from the SAMHSA Provider Survey and the SAMHSA Parent Survey. Both are retrospective pre-post measures that produce self-reported scores on change as a result of LAUNCH involvement with their service, not actual change scores. The Provider Survey asks for provider self-report on the extent of change in knowledge of children's mental and behavioral health and practices related to developmental assessment and use of mental health consultation. The Parent Survey asks about parent perceptions of the extent to which specific LAUNCH-supported programs have helped their parenting and their child's development. As it was worded, some grantees described the Parent Survey as a measure of parent satisfaction with the services, and not a parent outcome. As described in Chapter 1, the fact that these surveys only produce change scores, not separate baseline or posttest scores, means that they do not meet R-SEED standards for strength of evidence. We cannot calculate the standardized effect sizes required for the outcome analyses. Examination of the survey data is limited to descriptive analyses, the results of which are presented below.

6.1 SAMHSA Provider Survey

6.1.1 Survey Data

The local evaluators were expected to administer the Provider Survey each year of the grant, starting in year 2. Although it is possible that the same providers answered the survey across years, evaluators did not track the individual providers who responded each year. The evaluators used different data collection procedures. One grantee sent the survey to providers annually, using Survey Monkey. Other evaluators administered the surveys in person with the different provider groups. Evaluators were encouraged to administer the survey at the end of the program year or program session.

Retrospective pre-post outcomes were reported by home visitors in a majority of the home visiting programs (74%), by staff in a majority of the early childhood programs receiving mental health consultation (68%), by staff in nearly a third of the family strengthening programs (31%), and by health care providers in one program to integrate behavioral health in primary care (Exhibit 23). Grantees were not successful in obtaining high response rates to the Provider Survey. Across the program strategies, response rates ranged from 14 percent to 37 percent. These low response rates raise the possibility that the results may not accurately represent the outcomes among the full sample of providers in the programs. We cannot know whether the provider who failed to respond would have reported more or less change, compared to the responders. Therefore, in addition to the caveat concerning our inability to make valid causal inferences about the role of LAUNCH in producing outcomes for providers, we have to add the caveat that the results may not be accurate indicators of the amount of change actually experienced by providers.

Exhibit 23. Post Outcomes Reported on SAMHSA Provider and Parent Surveys Over All Implementation Years[a]

	Provider Survey	Parent Survey
Home Visiting		
# LAUNCH-supported programs across all grantees	31	26
# programs with provider data on self-reported change (% of all LAUNCH programs)	25 (81%)	11 (42%)
Average response rate on survey	14.5%	Not reported
Family Strengthening		
p# LAUNCH-supported programs across all grantees	61	62
# programs with provider data on self-reported change (% of all LAUNCH programs)	19 (31%)	13 (21%)
Average response rate on survey	28.1%	Not reported
Early Childhood Mental Health Consultation: Pre-K and Elementary Combined[a]		
# LAUNCH-supported programs across all grantees	33	33
# programs with provider data on self-reported change (% of all LAUNCH programs)	25 (76%)	1 (3%)
Average response rate	36.7%	Not reported
Integration of Behavioral Health in Primary Care		
# LAUNCH-supported programs across all grantees	25	11
# programs with provider data on self-reported change (% of all LAUNCH programs)	1 (4%)	3 (27%)
Average response rate on survey	29.7	Not reported

[a] Early childhood mental health consultation in preschool settings and school settings are combined because of small numbers of programs reporting on provider outcomes related to consultation in elementary schools.

Source: Provider outcomes reported by grantees in CSE Web Portal 2009- 2013; Parent outcomes reported in annual end-of-year local evaluation reports 2009 – 2013.

6.1.2 Findings on Provider Changes Related to LAUNCH

On average, the majority of providers in each of the program strategies reported that their knowledge of children's socio-emotional development, of appropriate service options for children with behavioral concerns, and their use of mental health consultation had increased "some" or "substantially" since the time when LAUNCH became involved with the program (Exhibit 24). In four of the program strategies, a majority of providers also reported an increase in the use of developmental screening in their practice settings. The one exception was in elementary schools, where fewer of the teachers who responded (45%) reported increased use of developmental screening.

Exhibit 24. Percentage of Providers Reporting Change in Knowledge and Practices as a Result of Project LAUNCH Involvement in their Program

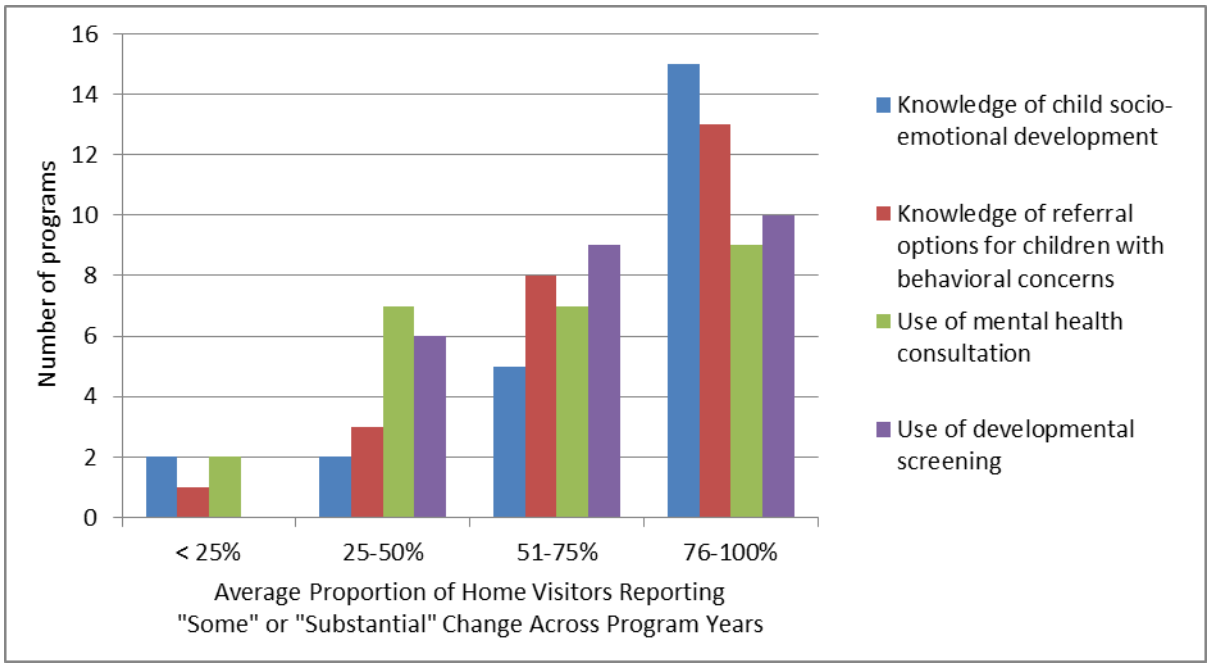
	Proportion of Providers Reporting "Some" or "Substantial" Change[a]			
	Knowledge of children's socio-emotional development and mental/behavioral health	Knowledge of options for follow-up services for children with mental/behavioral issues	Use of mental health consultation	Use of developmental screening in the practice setting
Integration of Mental/Behavioral Health in Home Visiting				
Mean	74.00	72.20	64.34	66.57
s.d.	26.48	23.30	30.41	25.71
Family Strengthening				
Mean	74.94	72.92	63.84	65.34
s.d.	22.02	21.43	24.14	28.38
Early Childhood Mental Health Consultation: Preschool				
Mean	84.21	68.85	72.84	70.54
s.d.	18.66	23.45	24.10	26.50
Early Childhood Mental Health Consultation: Grades K- 3				
Mean	76.06	77.69	68.28	44.97
s.d.	31.23	22.49	29.19	38.04
Integration of Behavioral Health in Primary Care				
Mean	66.84	72.39	65.49	73.55
s.d.	17.08	15.33	18.20	19.26

[a] If only providers who reported "Substantial" change were included, the percentage of providers would drop by more than half for the four survey items across all of the program strategies.

Source: Provider outcomes reported by grantees in CSE Web Portal 2009- 2013.

The large standard deviations in the data indicate that there is substantial variation in the average levels of change reported by providers in different programs. To illustrate this variation, Exhibit 25 shows the number of home visiting programs for which there was low, moderate or high percentages of providers reporting change on the survey items. (Similar exhibits for family strengthening and early childhood mental health consultation are presented in Appendix G. No exhibit is included for integration of behavioral health in primary care, since there was just one program that reported provider outcomes.) For the home visiting programs, the area in which home visitors were most likely to report change was knowledge of children's socio-emotional development: In 60 percent of the home visiting programs, between 76 and 100 percent of the home visitors reported increased knowledge. The area in which home visitors were least likely to report change was use of mental health consultation. In a third of the home visiting programs, most of the home visitors reported increased use.

Exhibit 25. Average Percentage of Home Visitors Reporting Change in Knowledge and Practices as a Result of Project LAUNCH Involvement in their Program *(n = 25 programs across 15 grantees reporting retrospective pre-post provider outcomes)*



Source: Provider outcomes reported by grantees in CSE Web Portal 2009- 2013.

6.2 SAMHSA Parent Survey

6.2.1 Survey Data

The local evaluators were expected to administer the Parent Survey each year of the grant, starting in year 2. It is possible that the same parents answered the survey across years; evaluators did not track the individual parents who responded each year. Across grantees, the evaluators also used different data collection procedures. Evaluators administered the surveys in person with the different parent groups at the end of the program year or program session. Across all reporting periods, responses on the Parent Survey were reported from over a third of home visiting programs, 20 percent of the family strengthening programs, and less than 10 percent of the early childhood programs and the primary care programs receiving LAUNCH-supported mental health consultation (Exhibit 26). This low response rate among these latter two types of programs could have been expected, since these are the two types of programs where parents are not the direct recipients of the LAUNCH services. This was borne out in the data.

6.2.2 Findings on Parent Satisfaction with LAUNCH Services

On average, parents were positive about the amount of help they received from the LAUNCH programs. Regardless of the type of program, a high rate of parents participating in the LAUNCH-supported home visiting programs felt that the program was helpful for their family, for their parenting skills, and for their child's development (Exhibit 26). The majority of the parents participating in family strengthening programs felt that the programs were helpful. Compared with

the parents in the home visiting programs, parents in the family support programs were similarly positive about the extent to which the programs helped their family or parenting; a smaller proportion of parents in family support programs gave high ratings on improving child outcomes, although the proportions were still very high (80% and higher).

Exhibit 26. Proportion of Parents Reporting that LAUNCH-Supported Program was "Somewhat" or "Very" Helpful

	Program Helps to Improve						
	Family and Parent Outcomes			Child Outcomes			
	Parenting skills	Overall family functioning	Ability to understand child's feelings and how to respond	Child's ability to express feelings in age appropriate way	Child's physical health	Child's expressive skills	Child's readiness for school/child's success in school/ preschool
Home Visiting (n = 11 programs, 491 parents)							
Average %	95.4%	97.5%	95.8%	87.8%	97.3%	92.5%	90.0%
S.D.	9.6	4.6	6.3	13.3	5.1	9.8	13.5
Family Strengthening (n = 13 programs, 333 parents)							
Average %	93.0%	95.3%	93.0%	83.8%	87.3%	88.6%	81.8%
S.D.	5.3	2.7	3.7	8.7	10.1	7.9	12.3
Early Childhood Mental Health Consultation (n = 1 programs, 55 parents)							
Average %	100%	96%	94%	99%	100%	97%	98%
S.D.	-	--	-	-	-	-	-
Integration of Behavioral Health in Primary Care (N =3 programs, 228 parents)							
Average %	91.5%	84.0%	79.0%	91.2%	76.3%	83.5%	74.4%
S.D.	3.5	18.3	21.2	8.5	25.4	19.1	14.8

Source: Parent outcomes reported in annual end-of-year local evaluation reports 2009 – 2013.

7. CONCLUSIONS AND DISCUSSION

7.1 What Did We Learn from the Cross-Site Evaluation?

This report contains some important information for SAMHSA about Project LAUNCH. The outcome data available at this point in Project LAUNCH indicate that the providers, parents, and children who are part of the LAUNCH initiative are on positive trajectories, even though the fact that the evaluation studies do not include comparison groups means that LAUNCH cannot claim to be the causal agent of this growth. The kind of evidence that supports statements about LAUNCH effectiveness will require that future evaluations test whether growth in program participants is greater than growth in similar families not in LAUNCH. Having this stronger evidence not only provides evidence to support the LAUNCH model, it also will allow for additional testing of key questions about implementation practices or systems partnerships that are the most effective for creating the kinds of changes LAUNCH was developed to address.

The outcome data also suggest that it takes time for grantees to implement effective programs. The Cohort 1 grantees found larger effects, relative to the other cohorts. This may indicate that rigorous evaluations would be most appropriate in the second half of the LAUNCH grants, after the grantees have had time to learn about how the programs are being implemented and to potentially institute program improvements, and time to design and implement more rigorous evaluations with appropriate comparison groups.

Further, Project LAUNCH grantees proved capable of implementing mental health enhancements to early childhood services and planning and, in some cases, initiating systems changes (see Volume I of this final report—Gwaltney, Goodson, Pfefferle, & Walker, 2014). Of the nearly 100 programs supported by Project LAUNCH, the Cross-Site Evaluation also found that the majority brought to the staff and participants a focus on mental and behavioral health, which was an over-riding objective of Project LAUNCH. To the extent that these LAUNCH-supported enhancements are sustained, the Project LAUNCH grantees may be able to effect lasting change in the extent to which the practices in the child service system in LAUNCH states and communities reflect an understanding of social-emotional development and mental health (Gwaltney, Goodson, Pfefferle, & Walker, 2014).

7.2 Evaluating Project LAUNCH: Challenges and Future Directions

The Project LAUNCH Cross-Site Evaluation was tasked with summarizing the findings in the local evaluations as well as providing technical assistance to grantees in designing meaningful evaluations. In the context of the focus of the local evaluations on grantee questions and grantee needs, of the relatively limited funding, and of the challenge of designing studies to understand the effects of enhancements rather than of whole programs, the technical assistance did not result in large-scale implementation of more rigorous local evaluations. The exception was the Special Studies, with whom the cross-site evaluators worked closely to support high quality research designs.

As a result of the evaluation context on the ground in the LAUNCH sites, the Cross-Site Evaluation did not have a rich database of locally-generated effects on which to build its meta-analysis. Obviously, some of the other designs that were considered, such as having the Cross-Site Evaluation team design and implement common evaluation designs in the LAUNCH sites, would have potentially generated more informative data. However, as discussed earlier, due to a number of critical concerns about the resources and skills of the local evaluators and the challenges of an intervention model built around enhancements versus new programs, the meta-analytic approach was judged to be the only feasible option at the beginning of the Cross-Site Evaluation.

The experiences of the LAUNCH Cross-Site Evaluation also make it necessary to reflect once again on the best design for the Cross-Site Evaluation. On the one hand, if LAUNCH is conceptualized as a multi-strategy intervention at the community level, intended to improve outcomes for all young children and families, the most appropriate evaluation design would be a design that compares outcomes for the LAUNCH community to similar outcomes in comparison communities, both before LAUNCH and during the LAUNCH grant period. This design depends on there being data on relevant outcomes at the community level, over time (prior to and after the LAUNCH grant begins), for the LAUNCH community and for potential comparison communities. Further, it must be recognized that a design with a single treatment unit (e.g., one LAUNCH community) will have very little power to detect an effect.

If instead, LAUNCH is conceptualized as an accumulation of programs, services, and systems activities, the most appropriate evaluation design would be a set of studies of the effects of individual LAUNCH services on providers and participating parents and children. Each LAUNCH project would then have to design and implement separate studies of each of the services being supported by LAUNCH. Rigorous studies would involve developing a valid design for each type of service, identifying an appropriate comparison group, and collecting data on both LAUNCH and comparison providers, parents, and children. For example, where LAUNCH funded enhancements of an existing program, the appropriate comparison group to estimate the effect of LAUNCH would be the same program but without the enhancements. Where LAUNCH funded a new program, the appropriate comparison group would be a sample of matched non-participants.

Both of these approaches require expertise in designing rigorous quasi-experimental studies, identifying extant data or collecting data on relevant outcomes, and conducting statistical analysis. All of these also require a sufficient budget for the full set of outcome evaluations. And, beyond these necessary skills and resources, the Project LAUNCH grantees would have needed strong incentives to commit to this level of evaluation, such as requirements for rigorous evaluations as a condition of receiving funding.

With limited funds and in the absence of program mandates related to conducting local-level comparison group studies, grantees implemented weaker designs and were uneven in their coverage of all of the LAUNCH-supported services and both adult and child outcomes. The result is that the findings from the cross-site outcomes evaluation have to be seen as preliminary and exploratory. The diverse grantee designs do not allow us to attribute definitively causal links between the observed overall changes in providers, parents, or children and LAUNCH.

The ambitiousness of the reach and objectives of Project LAUNCH need to be matched with ambitious standards for evaluation. Such alignment would better support SAMHSA's sustainability objectives for the Project LAUNCH grantee programs. As part of their funding, individual LAUNCH grantees have developed strategies for sustaining their activities in their community once the funding ends. However, some may find that the sustainability of Project LAUNCH as an initiative depends on rigorous evidence of its effect on children from well-designed and well-implemented evaluation studies. SAMHSA might consider providing grantees with stricter evaluation requirements and then holding grantees to these requirements.

7.3 Findings from the Cross-Site Evaluation of Outcomes

The evidence for estimating the effects of Project LAUNCH is limited by challenges to local evaluations conducting rigorous and comprehensive studies reflecting the broad programmatic variation among the grantees. Taken at face value, the average effects of LAUNCH strategies are moderate-to-small, with the smallest average effects on children. This report has made clear, however, that (a) nearly all of the results are generated by less rigorous studies; (b) less rigorous studies tend to have larger positive effect sizes; and (c) these study designs leave open the very real possibility that the changes observed in providers, parents, and children are caused by something other than LAUNCH—for example, history (other events happening in the lives of the respondents) or development (especially in the case of child outcomes).

The overall positive direction of the results is mirrored in the results from the SAMHSA Provider Survey and Parent Survey. On the Provider Survey, the results indicate that providers themselves attribute changes in their knowledge and behavior to the involvement of LAUNCH in their particular program. On the Parent Survey, the results indicate that most parents found the LAUNCH programs helpful in improving their parenting and their child's growth and development.

The examination of potential predictors of variation in effects focused on two implementation indices—one related to the depth of the integration of behavioral health in individual LAUNCH programs, and the second related to the breadth of initiatives to improve the child service system at the local and state levels. These analyses were generally inconclusive, although there were a few findings that suggested that these indices were related to the size of the observed effects.

7.4 Lessons for Project LAUNCH Going Forward

In spite of concerns about the quality of the evidence on LAUNCH effects, we can nevertheless discuss what the findings might tell us about LAUNCH going forward, should these results be confirmed in future, more rigorous research. First, the data suggest that outcomes for children were least often demonstrated in the local LAUNCH evaluations, versus outcomes for providers or parents. There are multiple hypotheses for why child outcomes were less frequently shown. First, the logic model for Project LAUNCH implies that child outcomes are primarily achieved through changes in the caretaking adults around the child, e.g., parents, child care providers, physicians, it may take longer for child outcomes to develop. Second, there may be methodological reasons why child outcomes were not shown, especially related to the use of measures that are either not valid

or not appropriate for the outcomes of interest. If child outcomes are of primary importance, it is possible that the key prevention and promotion strategies need to be expanded or revised to focus more strongly on direct services to children and on evaluation designs that account for the expected timing of the child outcomes.

Second, the relationships shown between the breadth of mental health integration activities and some outcomes suggest that how grantees approach mental health integration could make a difference. In particular, general staff training on topics related to child socio-emotional development and training on child assessment are not as strongly related to outcomes as are using mental health clinicians to work directly with staff; and, in the case of early childhood and school settings, focusing the mental health consultation on the broader classroom environment as opposed to individual children and families, may be beneficial. Outcomes reported earlier in Chapter 6, indicating that the majority of providers in each promotion and prevention strategy reported not only increased knowledge of socio-emotional development, but also knowledge about appropriate service options for children with behavioral concerns, hold promise for an important SAMHSA objective for young children—more timely access to effective prevention, early intervention, and care.

Third, the relationship of systems change activity to outcomes suggests that additional attention could be paid to the breadth and depth of grantees' efforts to bring about systems change at the state, tribal, and community levels. While services to children and families are usually considered the most direct route to improving the well-being of children and families, the findings suggest that a more effective service system and relevant policy development may provide a foundation for better outcomes.

All of these hypothesized lessons about Project LAUNCH need to be validated in another round of outcome analysis that includes more grantees and more years of implementation of the later cohorts of grantees. With additional research, we can, with more confidence, use the findings to validate the effectiveness of Project LAUNCH in improving outcomes for young children and their families and enhancing all levels of the early childhood system.

REFERENCES

Durlak, J.A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, 34, 917-928.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child development*,82(1), 405-432.

Fox, L., Dunpay, G., Hemmeter, M.L., Joseph, G., and Strain, P. (2010). Teaching Pyramid: A model for supporting social competence and preventing challenging behavior in young children. Center for the Socio-Emotional Foundations of Early Learning, Nashville, TN: Vanderbilt University.

Goodson, Barbara, Margaret K. Gwaltney, and Deborah Klein Walker. (2014). *Cross-site evaluation of Project LAUNCH: Interim findings.* OPRE Report 2014-46. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Goodson, B., Price, C., Darrow, C.,Parsad, P.,Williams, J. and Grindal, T. (in development). Rating System for Strong and Emerging Evidence Designs (R-SEED). Review protocol and evidence standards, Cambridge, MA: Abt Associates.

Gwaltney, M., Goodson, B., and Walker, D. (2013). *Cross-site evaluation of SAMHSA's Project LAUNCH Initiative: Key findings in the first year*. OPRE Report #2013-59. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Gwaltney, M., Goodson, B., Pfefferle, S., and Walker, D. (2014). *The legacy of SAMHSA's Project LAUNCH initiative: Findings from the cross-site evaluation*. Cambridge, MA: Abt Associates, Inc.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2007). Empirical benchmarks for interpreting effect sizes in research. Retrieved December 6, 2007, from http://www.mdrc.org/sites/default/files/full_84.pdf

Layzer, J. I., Goodson, B. D., Bernstein, L., & Price, C. (2001). *National Evaluation of Family Support Programs. Final Report Volume A: The Meta-Analysis*. Abt Associates Inc.

Layzer, J. I., Goodson, B. D., Bernstein, L., & Price, C. (2001). *National Evaluation of Family Support Programs. Final Report Volume A: The Meta-Analysis*. Abt Associates Inc.

Lipsey, Mark W. (1990). *Design sensitivity: Statistical power for experimental research* (Newbury Park, CA: Sage Publications).

Lloyd, C. M., Fishman, M., Bangser, M., & Rivera, A. (2013). *A First Look at the Head Start CARES Demonstration: Large Scale Implementation of Programs to Improve Children's Social-emotional Competence*. Office of Planning, Research and Evaluation, US Administration for Children and Families.

Nievar, M. A., Van Egeren, L. A., & Pollard, S. (2010). A meta-analysis of home visiting programs: Moderators of improvements in maternal behavior. *Infant mental health journal*, 31(5), 499-520.

Office of Special Education Programs (2014). *Positive Behaviorial Intervention and Supports*. OSEP Training Assistance Center, Washington, DC.

Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early childhood research quarterly*, 23(1), 10-26.

Substance Abuse and Mental Health Services Administration (2008). Cooperative Agreements for Linking Actions for Unmet Needs in Children's Health. Request for Applications (RFA) No. SM-08-011. Washington, DC: U.S. Department of Health and Human Services.

Substance Abuse and Mental Health Services Administration 2009). Cooperative Agreements for Linking Actions for Unmet Needs in Children's Health. Request for Applications (RFA) No. SM-09-009. Washington, DC: U.S. Department of Health and Human Services.

Substance Abuse and Mental Health Services Administration (2010). Cooperative Agreements for Linking Actions for Unmet Needs in Children's Health. Request for Applications (RFA) No. SM-10-012. Washington, DC: U.S. Department of Health and Human Services.

Sweet, M.A. and Appelbaum, M.I. (2004) Is Home Visiting an Effective Strategy? A Meta-Analytic Review of Home Visiting Programs for Families with Young Children? *Child Development*, 75(5), 1435-1456.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse (2011, November). Children Classified as Having an Emotional Disturbance intervention report: The Incredible Years. Retrieved from http://whatworks.ed.gov.

APPENDIX A: PROJECT LAUNCH GRANTEES

Grantee State	Grantee Agency	Project LAUNCH Community
Cohort 1		
Arizona	AZ Department of Health Services	Two zip codes in Phoenix
Maine	ME Department of Health and Human Services	Washington County
New Mexico	NM Department of Health	Santa Fe County
Rhode Island	RI Department of Health	City of Providence
Washington	WA State Department of Health	Yakima County
Wisconsin	Red Cliff Band of Lake Superior Chippewas	Red Cliff Reservation
Cohort 2		
California	CA Department of Public Health	East Oakland
District of Columbia	DC Department of Health	Wards 7 and 8
Illinois	IL Department of Human Services	Four communities on Chicago's West side: East/West Garfield Park; North/South Lawndale
Iowa	IA Department of Public Health	Seven zip codes in inner city Des Moines
Kansas	KS Department of Health and Environment	Finney County
Massachusetts	MA Department of Public Health	Boston
Michigan	MI Department of Community Health	Saginaw County
New York	NY Department of Health	Three communities in Westchester County: Yonkers, Ossining, and Port Chester
North Carolina	NC Division of Public Health	Guilford County
Ohio	OH Department of Health	Four counties of rural Appalachian Ohio: Athens, Hocking, Vinton, and Meigs
Oregon	OR Department of Human Services	Deschutes County
Wisconsin	WI Department of Health Services	Eight zip codes in Milwaukee
Cohort 3		
Colorado	North CO Health Alliance	Weld County
Connecticut	Wheeler Clinic, Inc.	New Britain
Missouri	Curators of the University of MO	Boone County
New York	Fund for Public Health in NYC	East Harlem and Hunts Point
Oregon	Multnomah Education Service District	Multnomah County
Texas	Aliviane, Inc.	Three census tracts within El Paso County and City

APPENDIX B: PROJECT LAUNCH EVALUATION REQUIREMENTS

EVALUATION REQUIREMENTS—2008, 2009, 2010 RFAs

2008 RFA (Cohort 1 grantees)

"You must devote the necessary resources to conduct grantee level evaluation of Project LAUNCH supported efforts at the State, Territorial, tribal and local level, to include assessment of: 1) the effectiveness of grant funded interventions; 2) the costs of implementing the program across the various populations served; 3) the quality and fidelity of implementation of evidence-based programs and practices (process evaluation); and 4) the strength of local and State partnerships. In addition, you must agree to meet the requirements of a national cross site evaluation including collection of common outcome measures and cost data." (p. 11)

2009 RFA (Cohort 2 grantees)

"Grantees are expected to design and implement comprehensive evaluations of their Project LAUNCH programs. Grantee-level evaluations should include process, outcome, and cost evaluation components. The outcomes component of the evaluation should aim to demonstrate potential linkages between project activities and improved outcomes both at the State/Territorial/Tribal and local levels, as identified in the LAUNCH logic models. Cost evaluation should include, at a minimum, costs of implementing the program (and individual program components) and cost per person served. In developing their evaluation plans, grantees may wish to consider outcome and process questions such as the following (findings may be included in the performance section of their Annual Progress Reports:

Outcome questions

- What were effects of the interventions on participants?
- What program/contextual factors were associated with outcomes?
- What individual factors were associated with outcomes?
- How durable were the effects?
- How sustainable were the programs?

Process questions

- How closely did implementation match the plan? What types of deviation from the plan occurred? What led to the deviations?
- What effect did the deviations have on the planned interventions and capacity to achieve desired outcomes?
- Who provided (program staff) what services (modality, type, intensity, duration), to whom (individual characteristics), in what context (system, community), and at what cost (facilities, personnel, dollars)?

2010 RFA (Cohort 3 grantees)

"Grantees are expected to follow a strategic process for carrying out their grants that is consistent with SAMHSA's Strategic Prevention Framework … built on a risk and protective factors approach to prevention, and requires communities to systematically:

- Assess their prevention and promotion needs based on epidemiological data,
- Build their prevention and promotion capacity,
- Develop a strategic plan,
- Implement effective community prevention programs, policies and practices, and
- Evaluate their efforts for outcomes

Grantees are expected to develop a Performance Assessment Plan to be submitted approximately 6 months after grant award. Performance Assessment Plan… includes both process and outcome components… You may also consider outcome and process questions, such as the following:

Outcome Questions:

- What was the effect of the intervention on key outcome goals?
- What program/contextual factors were associated with outcomes?
- What individual factors were associated with outcomes, including race/ethnicity?
- How durable were the effects?
- What were the effects of the project on key child, family and community-level outcomes?

As appropriate, describe how the data, including outcome data, will be analyzed by racial/ethnic group or other demographic factors to assure that appropriate populations are being served and that disparities in services and outcomes are minimized.

Process Questions:

- How closely did implementation match the plan?
- What types of changes were made to the originally proposed plan?
- What led to the changes in the original plan?
- What effect did the changes have on the planned intervention and performance assessment?
- Who provided (program staff) what services (modality, type, intensity, duration), to whom (individual characteristics), in what context (system, community), and at what cost (facilities, personnel, dollars)?

Grantees will be required to report on progress achieved, barriers encountered, and efforts to overcome these barriers in a performance assessment report to be submitted at least annually.

APPENDIX C: PROJECT LAUNCH SPECIAL STUDY DESIGNS

Site/ Study	Research Question	Outcome Measure	Sample/Sample Size	Design	Expected Date for Impact Results
1 Kansas—Finney County (Cohort 2)	What is the impact of the LAUNCH-supported child and family services on children's school readiness?	State-developed child development checklist (9 domains), administered pre-post during preschool year	At-risk 4-year-olds in state pre-kindergarten program Approximately 100 preschool children/year in programs in LAUNCH communities and 100 in programs in each non-LAUNCH community	Comparison of children in state pre-kindergarten program in LAUNCH counties vs. children in same program in 1-3 comparison counties. Study combines 4 cohorts of children (2010 – 2014). 5 years of program-level baseline measures (average child readiness scores for programs pre-LAUNCH)	Late fall, 2014
2: Massachusetts - Boston (Cohort 2)	What is the impact of the LAUNCH-supported services for families and children on the rate of reported cases of child maltreatment?	State Child Protective Services database	19 census tracts in LAUNCH community and a sample of matched census tracts outside of LAUNCH community	Comparison of rates of maltreatment over time in LAUNCH census tracts and non-LAUNCH census tracts. Short interrupted time sample with multiple baseline (pre-LAUNCH years) and multiple years during LAUNCH.	Late fall, 2014
3i California-East Oakland (Cohort 2)	What is the impact of the LAUNCH-supported services for families and children on kindergarten entry readiness scores?	Child assessment completed by kindergarten teachers as part of LAUNCH special study Parent survey on parent-child relationship, understanding of child development, parent mental health	200 kindergarten children in 9 elementary schools in the school district in the LAUNCH community	Comparison of average kindergarten readiness scores for children entering kindergarten from LAUNCH zip codes and children in same schools from non-LAUNCH zip codes. Baseline: fall 2011; LAUNCH: fall 2013	Late fall, 2013

Site/ Study	Research Question	Outcome Measure	Sample/Sample Size	Design	Expected Date for Impact Results
3ii California-East Oakland (Cohort 2)	What is the impact of the LAUNCH-supported services for families and children on student achievement in grade 2?	Grade 2 CA Standards Test (CST) For ELs, CA English Language Development Test (CELDT) % students with identified special needs, % students suspended, % retained in grade	2nd grade students I 9 elementary schools (sample size not known)	Comparison of average achievement scores for children from LAUNCH zip codes and children in same schools from non-LAUNCH zip codes. Baseline: 2010- 2011; LAUNCH: 2013 - 2014	Late fall, 2014
4i Wisconsin-Milwaukee (Cohort 2)	What is the impact of the LAUNCH-supported services for families and children on readiness at kindergarten entry?	District administered kindergarten assessment: Measures of Academic Progress (MAP) (literacy & numeracy)	TBD	Comparison of trend lines for children from LAUNCH zip codes and children from non-LAUNCH zip codes in the same elementary schools across pre-LAUNCH and LAUNCH years. Baseline: 2011; LAUNCH: 2012-2014	Spring, 2015
4ii Wisconsin-Milwaukee (Cohort 2)	What is the impact of the LAUNCH-supported services for families and children on children's academic outcomes in grade 1?	District administered kindergarten assessment: Measures of Academic Progress (MAP) (literacy & numeracy), attendance, suspensions	TBD	Comparison of outcomes for children from LAUNCH zip codes and children from non-LAUNCH zip codes in the same elementary schools across pre-LAUNCH and LAUNCH years. Baseline: 2011; LAUNCH: 2012-2014	Spring, 2015
5i Red Cliff (Cohort 1)	What is the impact of the LAUNCH-supported services for families and children on the developmental status and school readiness of children at ages 1 - 5?	Assessments administered by early childhood program: Preschool: PPVT, Social Skills Rating System, Boehm-3 Preschool I/T: Bayley Scales	Between 32 and 50 children assessed in annual cohort. (Sample represents > 90% of children in this age group in the tribal community at each time point.)	Time lag design comparing children's developmental status pre-LAUNCH and during LAUNCH. Baseline: 2005-06: LAUNCH: 2006 – 2013 Developmental data for 4 & 5 year olds will be augmented with data from 1-3 year olds beginning in 2009/2010.	Spring, 2014

Site/ Study	Research Question	Outcome Measure	Sample/Sample Size	Design	Expected Date for Impact Results
5ii Red Cliff (Cohort 1)	What is the impact of the LAUNCH-supported services for families, children and schools on student academic outcomes in grades K – 3 (ages 6 – 8 years)?	District data on grades, state proficiency test (grade 3), attendance, special needs	Sample size TBD. Sample will represent > 90% of children in this age group in the tribal community	Time lag design comparing children's academic outcomes pre-LAUNCH and during LAUNCH Baseline: 2005-06: LAUNCH: 2006 – 2013	Spring, 2014
6i Maine-Washington County (Cohort 1)	What is the impact of the LAUNCH-supported early intervention services for families and children on the birth outcomes of newborns and their mothers?	Birth weight (low and very low birth weight babies) and other birth outcomes from state databases	LAUNCH county and matched comparison county	Short interrupted time sample following trends over 3 years pre-LAUNCH and 4 years of LAUNCH Baseline: 2006 – 2008; LAUNCH: 2009 - 2013	Late fall, 2013
6ii Maine-Washington County (Cohort 1)	What is the impact of LAUNCH-supported early intervention services on the health outcomes of babies born to opiate dependent, on the well-being and perceptions of the mothers, and on use and costs of acute health care?	Primary data collection involving interviews with mothers multiple times pre- and post-natally	Opiate-dependent mothers who receive LAUNCH support services pre- and post-natally and similar mother who experience the current standard of care	Comparison of outcomes for two groups of mothers during 2012-13	Spring, 2014

APPENDIX D: SAMHSA EXPECTATIONS FOR MEASUREMENT OF OUTCOMES IN LOCAL EVALUATIONS (2011)

	Cohort 1	Cohort 2	Cohort 3
Systems outcomes			
Level 1: systems surveys in portal each October, systems level GPRA measures only each April	Required at state/tribal and community levels	Required at state/tribal and community levels	Required at community level
Level 2: longitudinal measurement of systems change/infrastructure development activities and sustainability efforts	Required at state/tribal and community levels	Required at state/tribal and community levels	Required at community level
Level 3: baseline and annual measurement of collaboration using an established survey instrument (e.g., Wilder Collaboration Factors Inventory, PARTNER social network analysis tool)	Not Required or Recommended	Recommended—annual measurement, assessing change over time	Required—annual measurement, assessing change over time
Provider outcomes			
Level 1: annual provider survey (SAMHSA items)	Required for *all* LAUNCH-supported service strategies	Required for *all* LAUNCH-supported service strategies	Required for *all* LAUNCH-supported service strategies
Level 2: pre-post measurement, expanded provider outcome measures	Recommended for 1-3 key strategies	Required for 1-3 key strategies	Required for all LAUNCH-supported service strategies
Level 3: QED with comparison group of non-LAUNCH providers, expanded provider outcome measures	Recommended for 1-3 key strategies	Recommended for 1 -3 key strategies	Recommended for 1 -3 key strategies
Parent outcomes			
Level 1: annual parent survey (SAMHSA items)	Required for *all* LAUNCH-supported service strategies that involve direct work with families	Required for *all* LAUNCH-supported service strategies that involve direct work with families	Required for *all* LAUNCH-supported service strategies that involve direct work with families
Level 2: pre-post measurement, expanded parent outcome measures	Recommended for 1-3 key strategies that involve direct work with families	Required for 1-3 key strategies that involve direct work with families	Required for all LAUNCH-supported service strategies that involve direct work with families
Level 3: QED with comparison group of non-LAUNCH parents, expanded parent outcome measures	Recommended for 1-3 key strategies that involve direct work with families	Recommended for 1 -3 key strategies that involve direct work with families	Recommended for 1 -3 key strategies that involve direct work with families
Child outcomes			
Level 1: annual child outcomes	Recommended for 1-3 key strategies that involve direct work with children	Recommended for 1-3 key strategies that involve direct work with children	Recommended for 1-3 key strategies that involve direct work with children
Level 2: pre-post measurement of child expanded child outcome measures	Recommended for 1-3 key strategies that involve direct work with children	Recommended for 1-3 key strategies that involve direct work with children	Recommended for 1-3 key strategies that involve direct work with children
Level 3: QED with comparison group of non-LAUNCH children, expanded child outcome measures	As part of Special Study	As part of Special Study	As part of Special Study

APPENDIX E: SAMHSA PROVIDER AND PARENT SURVEYS

Provider Survey

As a result of the involvement of Project LAUNCH in this program, what is the extent of change in the following (*no change, a little change, some change, substantial change*):

1. Knowledge of children's socio-emotional and behavioral health and development
2. Knowledge of available options for follow-up services for children with mental or behavioral health issues
3. Use of mental health consultation for children with mental or behavioral health issues
4. Use of screening/assessment of children in work setting

Parent Survey

1. Has *[specific service or program]* offered support that has helped your family?

1	*2*	*3*	*4*
Not at all			*A lot*

2. Has *[specific service or program]* helped you to be a better parent to your child/children?

1	*2*	*3*	*4*
Not at all			*A lot*

3. Has *[specific service or program]* helped you to better understand what your child is feeling and how to respond?

1	*2*	*3*	*4*
Not at all			*A lot*

4. Has *[specific service or program]* helped your child's behavior?

1	*2*	*3*	*4*
Not at all			*A lot*

5. Has *[specific service or program]* helped your child to be physically healthy?

1	*2*	*3*	*4*
Not at all			*A lot*

6. Has *[specific service or program]* helped your child to express his/her feelings in a positive/age appropriate way?

1	*2*	*3*	*4*
Not at all			*A lot*

7. Has *[specific service or program]* helped your child to be ready for school or successful in school?

1	*2*	*3*	*4*
Not at all			*A lot*

APPENDIX F: R-SEED® STANDARDS FOR STRENGTH OF EVIDENCE

Excerpts from Rating System for Strong and Emerging Evidence Designs (R-SEED®)[23]:

Standards for Randomized Control Trials								
R-SEED® Criterion	**Highest Possible R-SEED® Evidence Rating**							
	Strong	**Emerging**			**Limited**			**Weak**
Outcome Standards:								
Face Validity	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reliability	Yes		Yes	Yes	Yes			
Not Over-aligned	Yes	Yes		Yes		Yes		
Consistency of Measurement	Yes	Yes	Yes				Yes	
Attrition Standards								
Attrition Standards Met	Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]
Integrity of Randomization								
Randomization not undermined	True[b]	True[b]	True[b]	True[b]	True[b]	True[b]	True[b]	True[b]

Quasi-Experimental Designs					
R-SEED® Criterion	**Highest Possible R-SEED® Evidence Rating**				
	Strong	**Emerging**		**Limited**	**Weak**
Outcome Standards:					
Face Validity	Yes	Yes	Yes	Yes	Yes
Reliability	Yes	Yes	Yes	Yes	
Not Over-aligned	Yes	Yes	Yes		
Consistency of Measurement Standards Met	Yes	Yes	Yes	Yes	
Pre-Test Standards					
Baseline Equivalence Standards Met	Yes	Yes			

[23] Goodson, B.D., Price, C., Darrow, C., Parsad, A. & Williams, J. (In development). *Rating System for Strong and Emerging Evidence Designs (R-SEED®). Review Protocol and Evidence Standards*. Cambridge, MA: Abt Associates Inc.

Standards for Randomized Control Trials								
R-SEED® Criterion	**Highest Possible R-SEED® Evidence Rating**							
	Strong	**Emerging**			**Limited**		**Weak**	
Quasi-Experimental Designs (continued)								
Baseline Equivalence Standards Not Met, but Model Controls for Pre-test			Yes		Yes		Yes	
Pre-test / Post-test Correlation is:	High	Moderate	High		Moderate			
R-SEED® Criterion	**Highest Possible R-SEED® Evidence Rating**							
	Strong	**Emerging**			**Limited**		**Weak**	
Design Confounds								
Free of Substantially Different Characteristics Confound	Yes	Yes	Yes					
Free of n=1 Person Provider Confound	Yes	Yes	Yes		Yes			
Free of n=1 Administrative Unit Provider Confound	Yes							
Free of Off-year Comparison Group Confound	Yes							
Pre-Post								
R-SEED® Criterion	**Highest Possible R-SEED® Evidence Rating**							
	Strong	**Emerging**	**Limited**				**Weak**	
Outcome Standards:								
Face Validity		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reliability		Yes	Yes	Yes	Yes	Yes		Yes
Not Overaligned		Yes	Yes	Yes	Yes	Yes		
Consistency of Measurement Standards								
The same measures is used at all pre and post time points		Yes	Yes	Yes	Yes	Yes	Yes	No

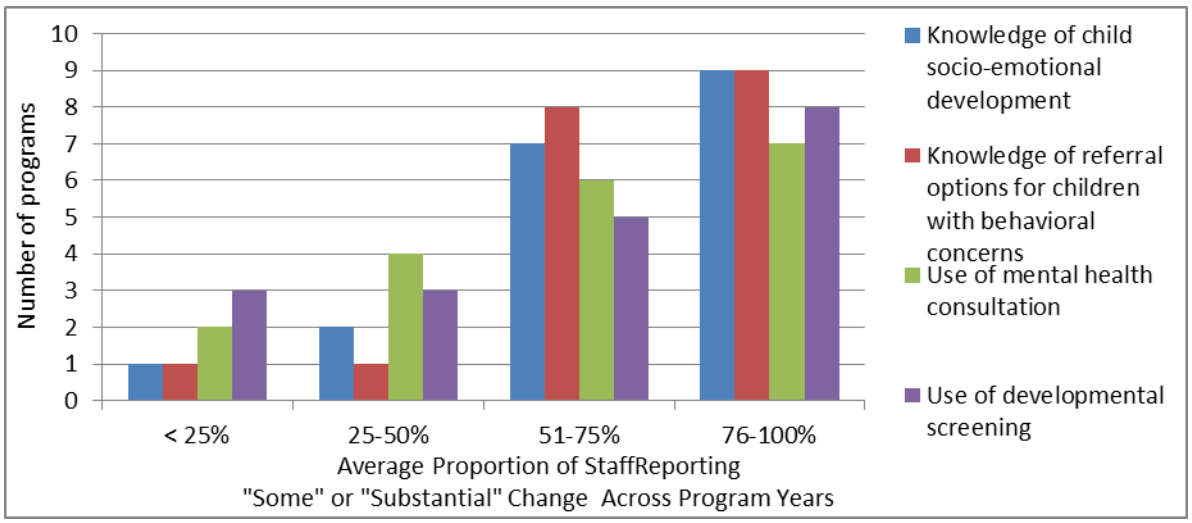
Standards for Randomized Control Trials								
R-SEED® Criterion	**Highest Possible R-SEED® Evidence Rating**							
	Strong	**Emerging**			**Limited**		**Weak**	
Instrument for outcome measure changes during study, but outcomes are normalized via z-scoring using population means and standard deviations		No	No	No	No	No		Yes
Pre-Post Designs (continued)								
Similar data collectors and data collection modes at pre and post time points		Yes	Yes	Yes	Yes	Yes		Yes
Design Confounds								
Free of Substantially Different Characteristics Confound		Yes						
Free of n=1 Person Provider Confound		Yes	Yes	Yes	Yes	Yes		
R-SEED® Criterion	**Highest Possible R-SEED® Evidence Rating**							
	Strong	**Emerging**	**Limited**				**Weak**	
Pre-tests at 3+ time points *and* Baseline projection demonstrated and accounted for w/ appropriate model *and* Counterfactual demonstrated by comparing pre-post gain on target outcome with pre-post gain on a non-equivalent dependent measure		Yes						
Pre-tests at 3+ time points *and* Baseline projection demonstrated accounted for w/ appropriate model			Yes					

Standards for Randomized Control Trials								
R-SEED® Criterion	**Highest Possible R-SEED® Evidence Rating**							
	Strong	**Emerging**			**Limited**		**Weak**	
Pre-Post Designs (continued)								
Pre-tests at one or two time points *and* Counterfactual demonstrated by comparing pre-post gain on target outcome with pre-post gain on a non-equivalent dependent measure				Yes				
Counterfactual for expected pre-post gain in the absence of treatment demonstrated by comparing pre-post gain to norms					Yes			
Treatment removal design with at least: one measurement taken before treatment, one measurement taken during or at end of treatment, and one measurement taken after treatment has been removed.						Yes		

APPENDIX G. FINDINGS ON THE SAMHSA PROVIDER SURVEY: PROGRAM VARIATION IN RESPONSES BY PROGRAM STRATEGY

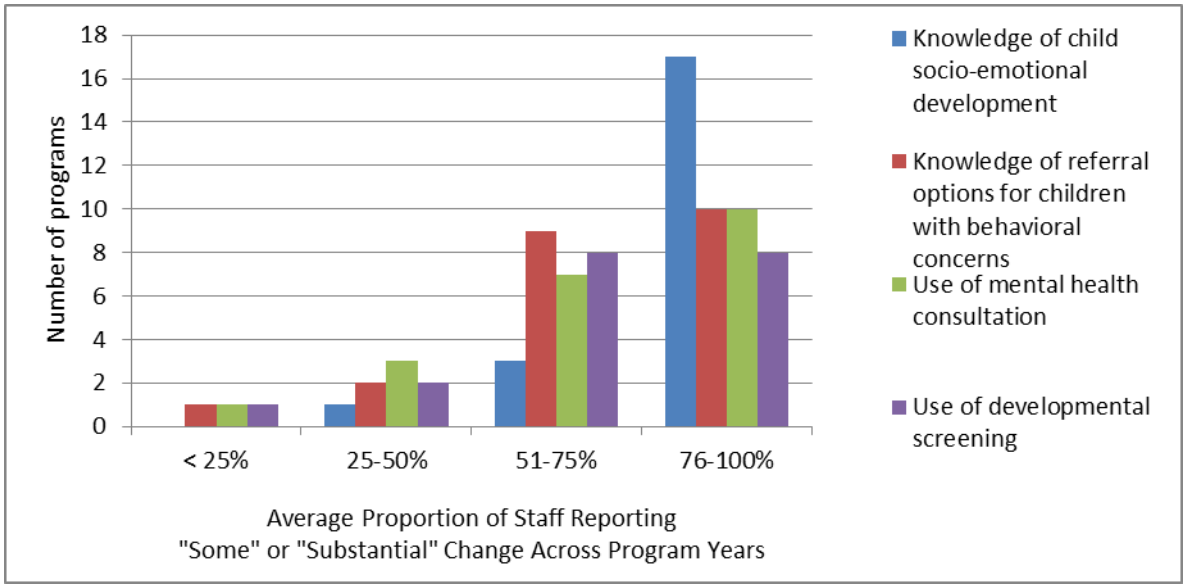
Family Strengthening Programs

Average Percentage of Staff in Family Strengthening Programs Reporting Change in Knowledge and Practices as a Result of Project LAUNCH Involvement *(n = 19 programs across 12 grantees reporting retrospective pre-post provider outcomes)*



Early Childhood Mental Health Consultation: Preschool

Average Percentage of Staff in Early Childhood Education and Care Programs Receiving Mental Health Consultation Reporting Change in Knowledge and Practices as a Result of Project LAUNCH Involvement *(n = 21 programs across 16 grantees reporting retrospective pre-post provider outcomes)*



Mental Health Consultation in Early Childhood: Kindergarten – Grade 3

Average Percentage of Staff in Elementary Schools Receiving Mental Health Consultation Reporting" Change in Knowledge and Practices as a Result of Project LAUNCH Involvement *(n = 9 programs across 7 grantees reporting retrospective pre-post provider outcomes)*